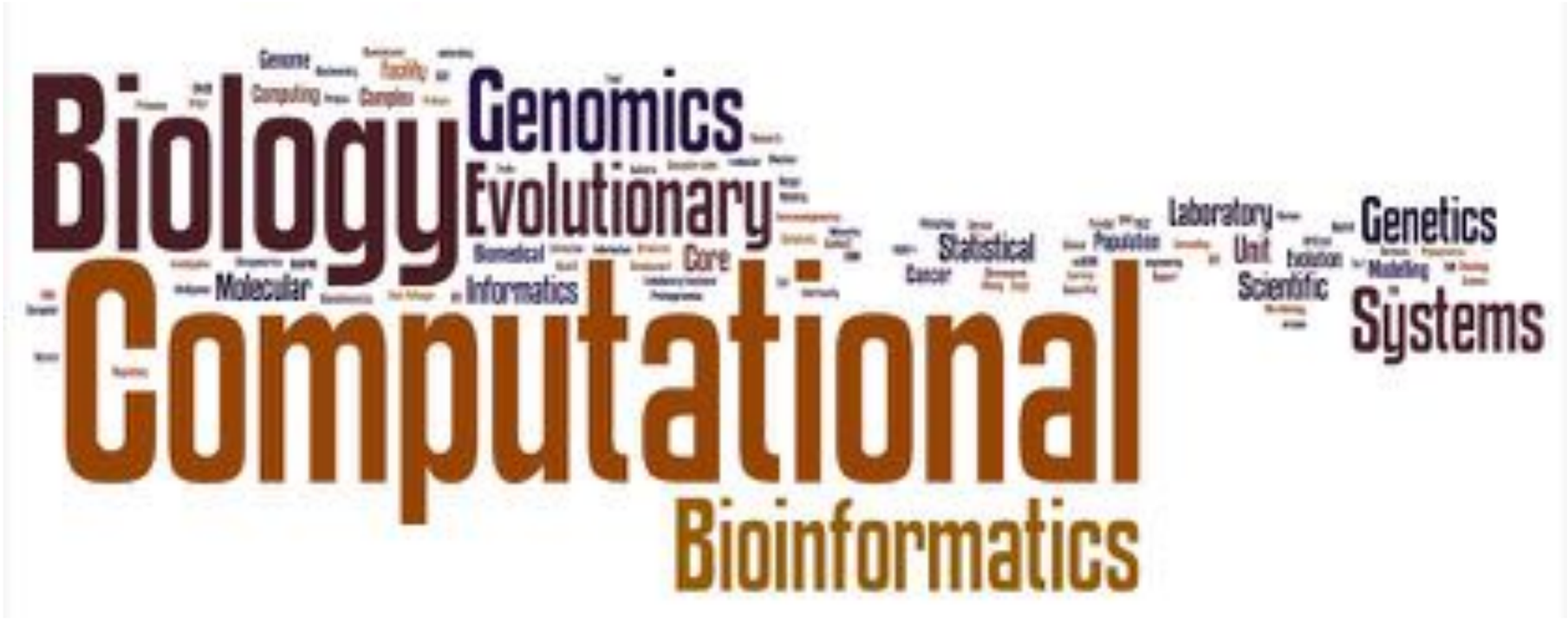
Молекулярное моделирование в биологии

к.ф.-м.н. Алексей Константинович Шайтан

alex@intbio.org
http://intbio.org

Вычислительная биология, биоинформатика и молекулярное моделирование, вычислительная химия



- Computational biology
- Bioinformatics
- Computational chemistry
- (Bio)Molecular modeling

2013 Chemistry Prize

Taking the Experiment to Cyberspace

The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel *"for the development of multiscale models for complex chemical systems"*.

- Press release
- Popular information
- Advanced information

Photo: © Harvard University

Martin Karplus

Martin Karplus, U.S. and Austrian citizen. Born 1930 in Vienna, Austria. Ph.D. 1953 from California Institute of Technology, CA, USA. Professeur Conventionné, Université de Strasbourg, France and Theodore William Richards Professor of Chemistry, Emeritus, Harvard University, Cambridge, MA, USA.

- More on Martin Karplus

Photo: S. Fisch

Michael Levitt

Michael Levitt, U.S., British and Israeli citizen. Born 1947 in Pretoria, South Africa. Ph.D. 1971 from University of Cambridge, UK. Robert W. and Vivian K. Cahill Professor in Cancer Research, Stanford University School of Medicine, Stanford, CA, USA.

- Have a look at Michael Levitt's photo gallery

Photo: Wikimedia Commons

Arieh Warshel

Arieh Warshel, U.S. and Israeli citizen. Born 1940 in Kibbutz Sde-Nahum, Israel. Ph.D. 1969 from Weizmann Institute of Science, Rehovot, Israel. Distinguished Professor, University of Southern California, Los Angeles, CA, USA.

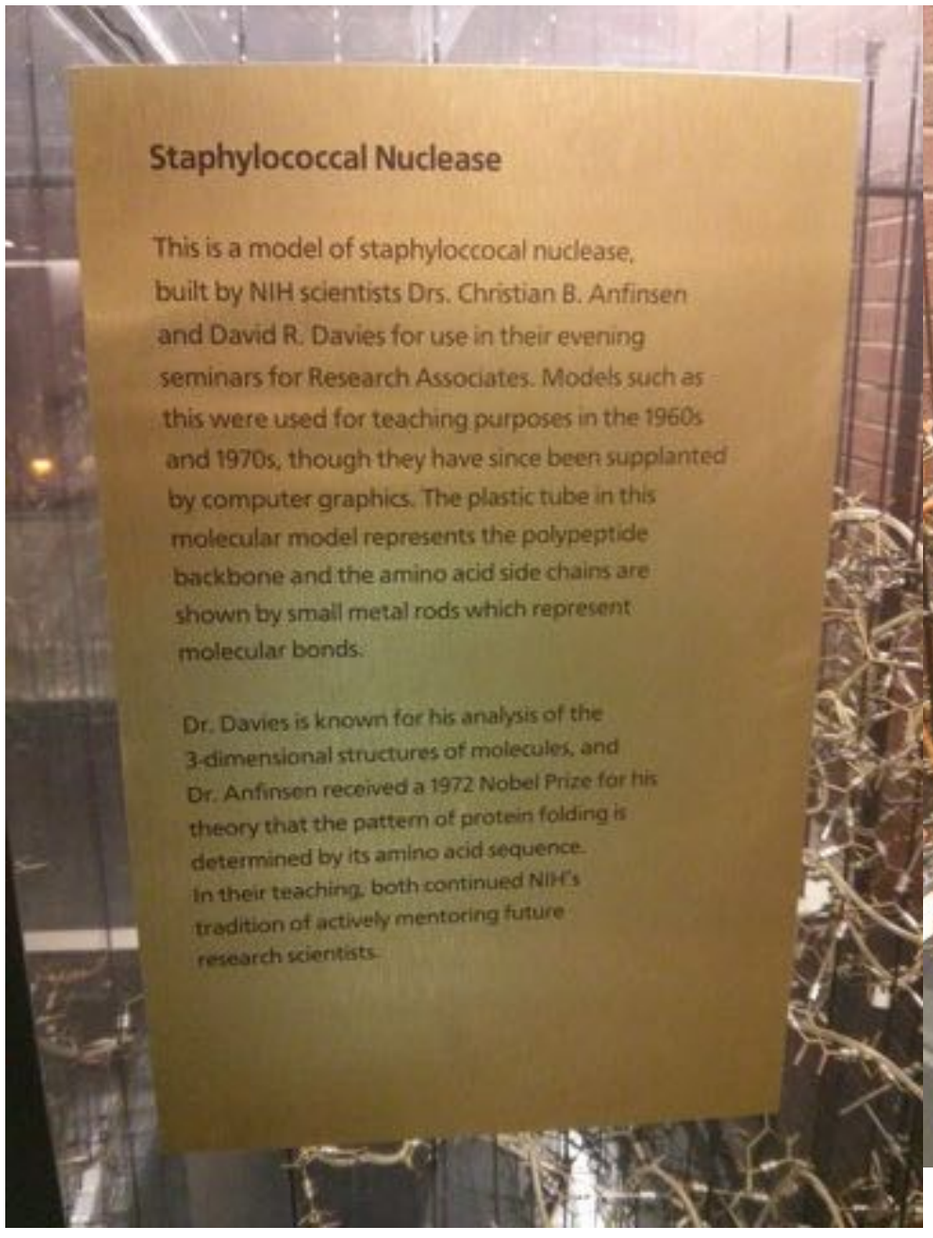
- Interviews with Chemistry Laureate Arieh Warshel

Молекулярное моделирование

- Построение и изучение моделей молекул с целью понимания их строения, функций, возможности рационального дизайна молекул с заданными свойствами.

«Модель есть абстрактное представление реальности в какой-либо форме (например, в математической, физической, символической, графической или дескриптивной), предназначенное для представления определённых аспектов этой реальности и позволяющее получить ответы на изучаемые вопросы»

Modeling in pre-computer era …



Молекулярное моделирование

- Molecular modeling
- Computer simulations
- Molecular simulations

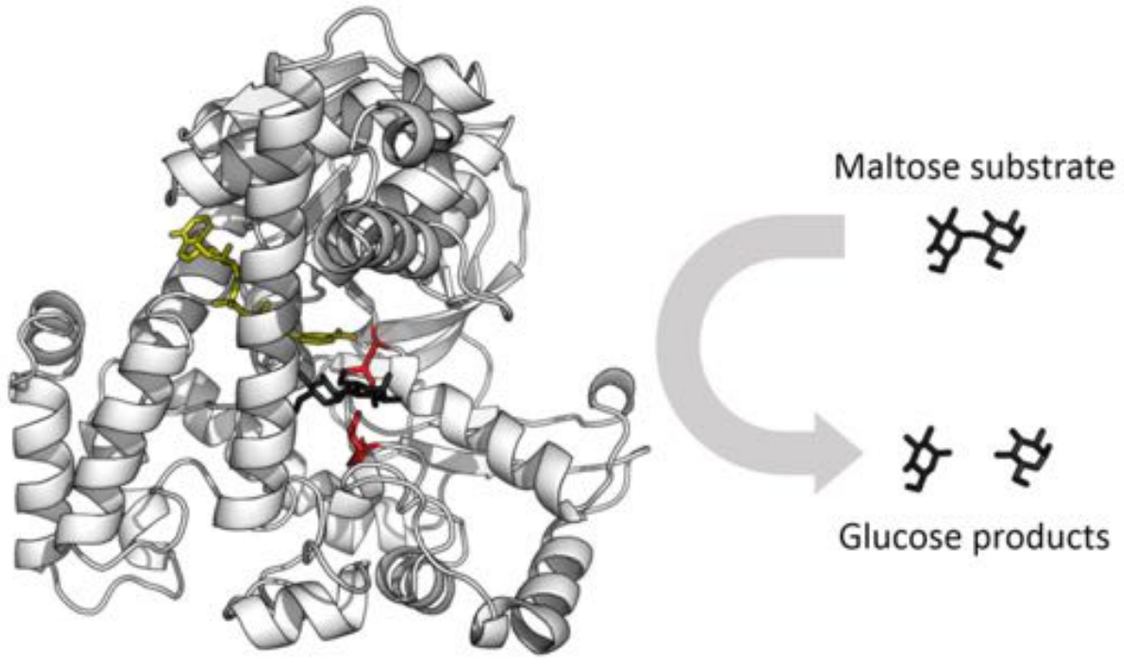
Modeling and simulation (**M&S**) at simple terms is a substitute for physical experimentation, in which computers are used to compute the results of some physical phenomenon.

Computer simulation is the reproduction of the behavior of a system using a computer to simulate the outcomes of a mathematical model associated with said system. Since they allow to check the reliability of chosen mathematical models, computer simulations have become a useful tool for the mathematical modeling of many natural systems in physics (computational physics), astrophysics, climatology, chemistry, biology and manufacturing, human systems in economics, psychology, social science, health care and engineering.

Молекулярное моделирование

Виды моделей:

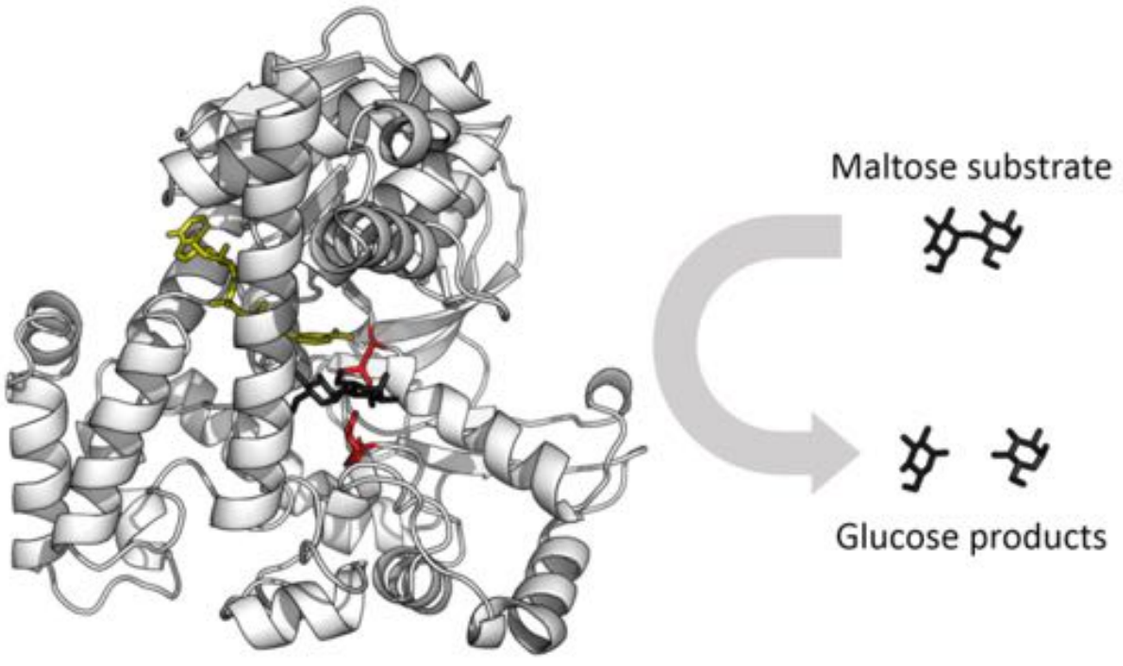
- Структурная – позиции атомов в пространстве.
- Динамическая
- Механическая
- Квантовая
- **Физическая** – учитывает взаимодействия между атомами.
- Компьютерная



Место молекулярного моделирования в исследованиях

- Эксперимент не дает нам прямой информации о положении атомов – все наши представления о строении биомолекул – результат построения (компьютерных) моделей, зачастую с использованием *a priori* информации о физических взаимодействиях между атомами.

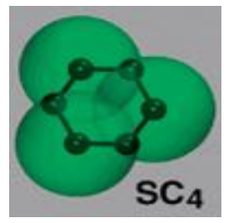
- Зная физические/эмпирические законы взаимодействия молекул мы можем попробовать предсказать их строение и поведение.

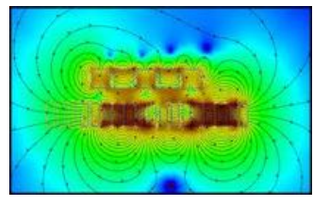


Modeling at various scales

$$\hat{H}\Psi = E\Psi$$

$$\vec{F} = m\vec{a}$$





Quantum chemistry	Molecular mechanics	Coarse grain models	Continuum models

Model accuracy

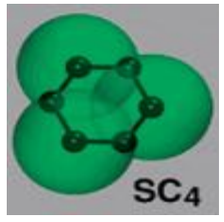
Computational complexity

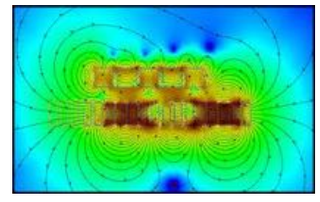
Time scales under study

System sizes under study

“Static” vs “dynamic” modeling

$$\hat{H}\Psi = E\Psi \qquad \vec{F} = m\vec{a}$$





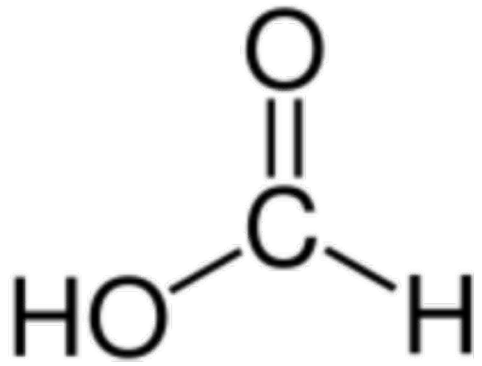
	Quantum chemistry	Molecular mechanics	Coarse grain models	Continuum models
Static	Electronic structure calculations, Structure optimization, Spectral calculations	Structure optimization, Vibration spectra calculation		Mechanical properties
Dynamic	Ab initio molecular dynamics, Car-Parrinello Molecular Dynamics	Molecular dynamics	Coarse grain molecular dynamics	Hydrodynamics, aerodynamics, heat transfer, diffusion, etc.

Dynamics: how things change with time?

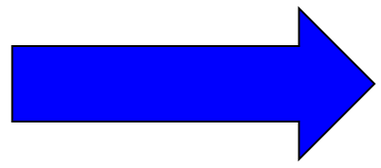
MD basic idea in a nutshell

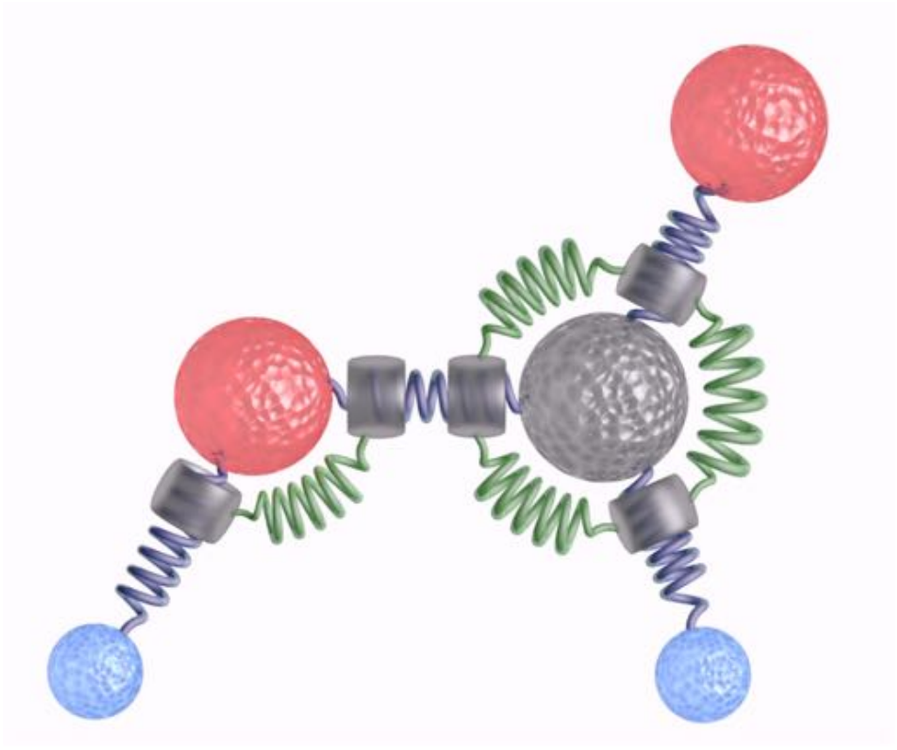
- Use classical mechanics to represent atoms and their interactions

$\vec{F} = m * \vec{a}$ - Newton's second law links force and acceleration



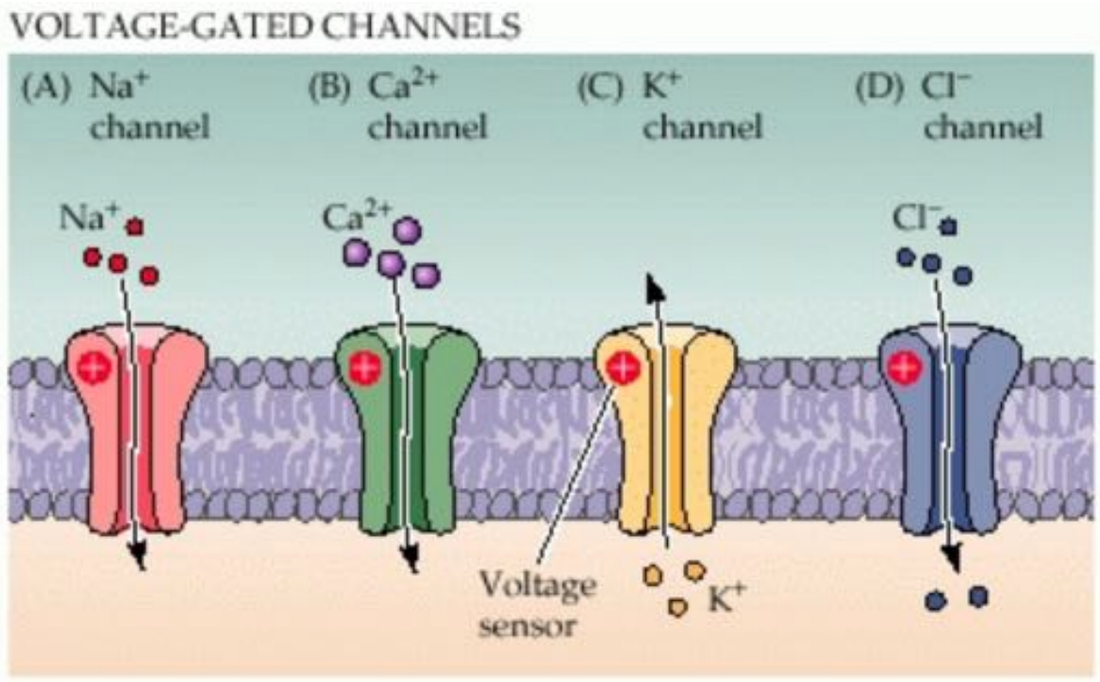
Formic acid





Example: switching of an ion channel

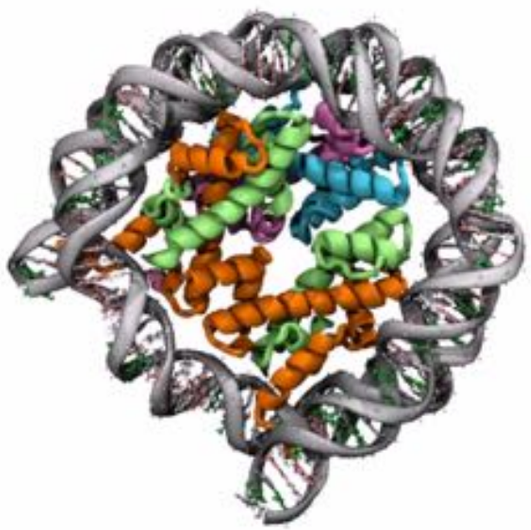
Science 2012, Morten Ø. Jensen et al.
“Mechanism of Voltage Gating in Potassium Channels”



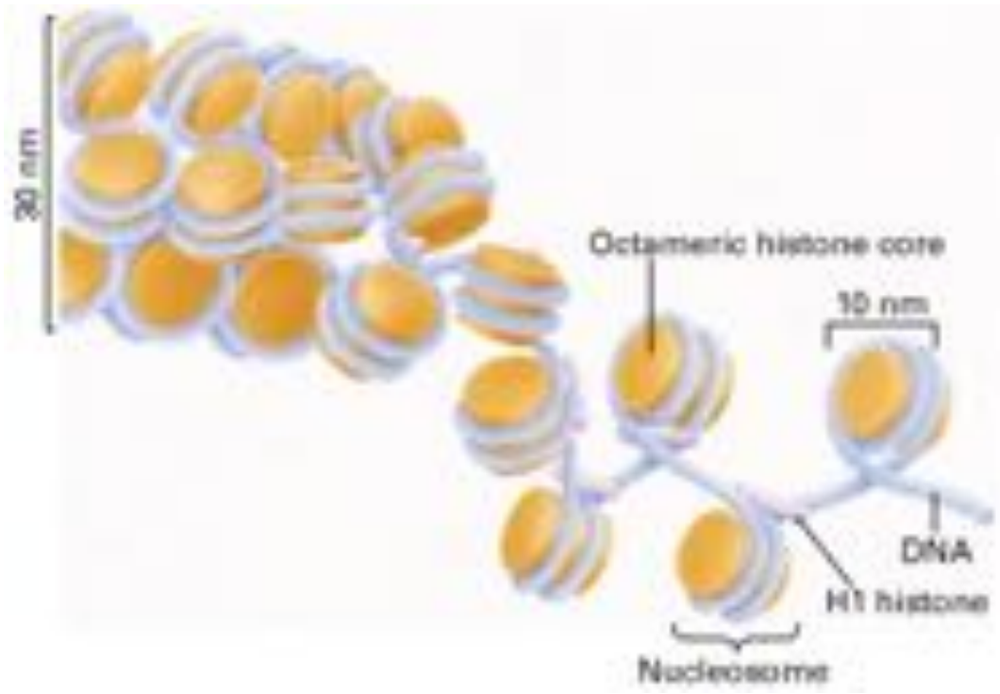
Neuroscience. 2nd edition.
Purves D, Augustine GJ, Fitzpatrick D, et al., editors.
Sunderland (MA)

Potassium voltage gated channel
Simulation details:
~250 μs
~200 000 atoms

Example 2: DNA packing in nucleosomes



Unwrapping of DNA from nucleosome



Solenoid model of the 30-nm condensed chromatin fiber in a side view

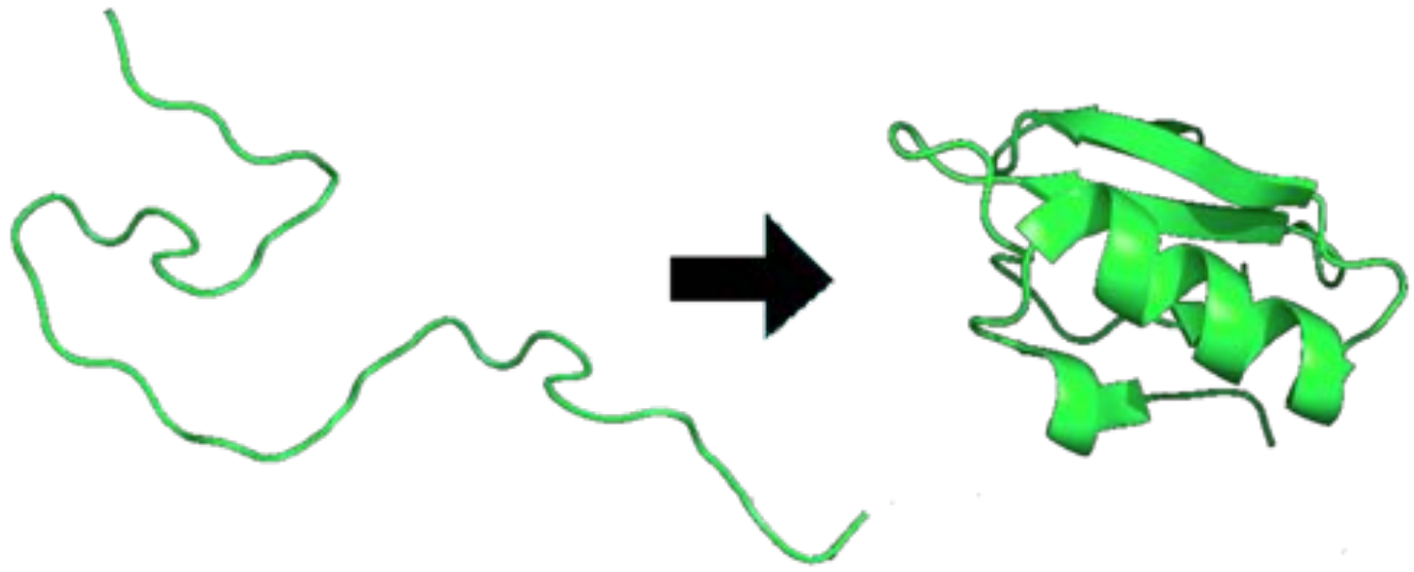
Molecular Cell Biology. 4th edition.
Lodish H, Berk A, Zipursky SL, et al.
New York: W. H. Freeman; 2000.

Simulation details:
~300 000 atoms

When and why use MD?

- MD as “computational microscope”
- Ultimate, but tough goal :

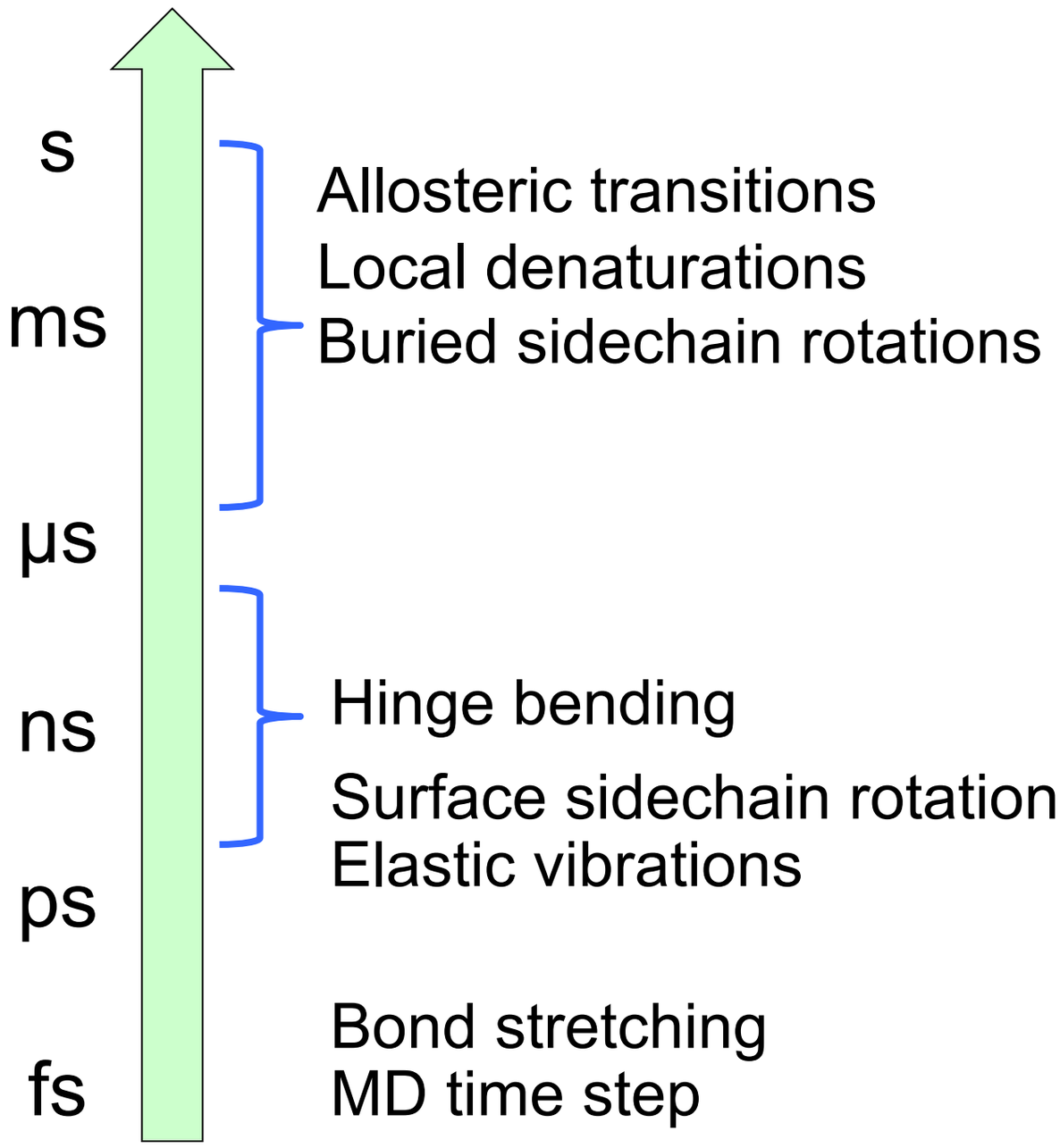
given the sequence -> simulate protein folding -> simulate function – not practically feasible!



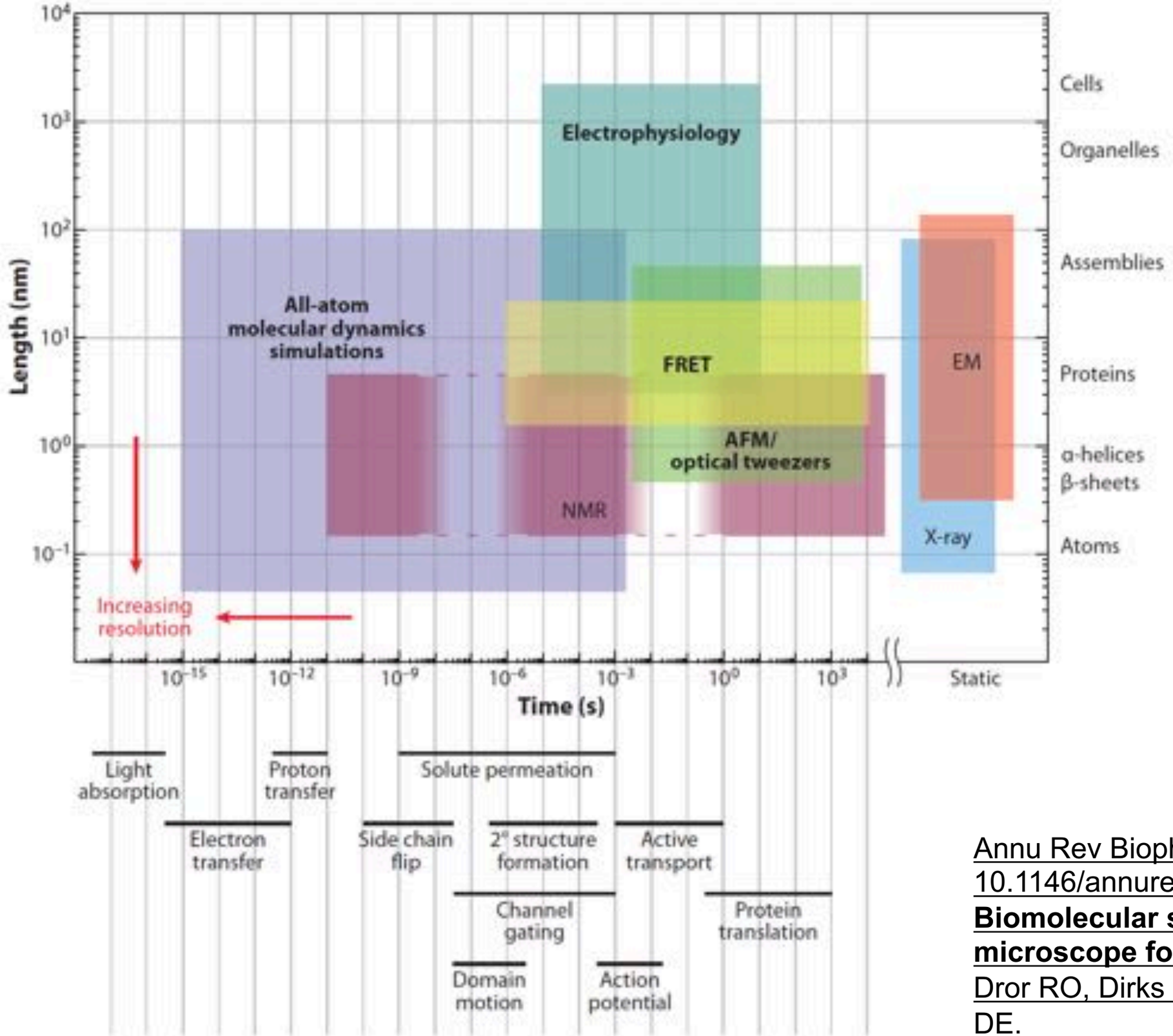
- Need experimental 3D structure of protein to start with.
- Where to get it? => experiment or homology modeling

Time scales of protein motions

Simulation time, approx.	MD steps
Inf.	10^{15}
250 years	10^{12}
100 days	10^{9}
0.1 day	10^{6}
	10^{3}
	1



When and why use MD?



Annu Rev Biophys. 2012;41:429-52. doi: 10.1146/annurev-biophys-042910-155245.
Biomolecular simulation: a computational microscope for molecular biology.
Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE.

Basic theory behind MD simulations

- Physical laws and approximations
- Classical force fields
- Numerical integration, integration step and timescales

From quantum to classical

Feasible for 3 atoms

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi$$

Time dependent non-relativistic Schrödinger equation

=>

$$\Psi_{\text{total}} = \psi_{\text{electronic}} \times \psi_{\text{nuclear}}$$

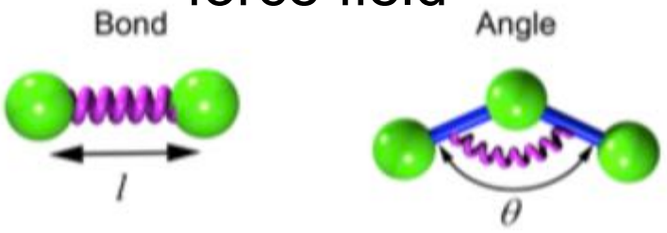
Born-Oppenheimer approx., Separate nuclei motion from electronic

(not valid for photochemical reactions, charge transfer)

||
V

Feasible for millions to billions of atoms

Construct empirical approximate force field



<=

Assume classical motion of nuclei on electronic potential energy surface

$$\vec{F} = m * \vec{a}$$

Newton's second law

Usually very simply, no bond breaking, ground state, pairwise interactions

Not true for motions faster than 100 fs (bond, angle vibrations), no tunneling, no proton hopping

What to expect/not expect from MD?

No, per se:

1) Chemical reactions
2) Protonation/deprotonation
3) Charge transfer
4) Excited states

Although possible with various hybrid methods: ab initio MD, QM/MM, etc.

Yes, if sufficient time and force field accuracy:

1) Conformational motions
2) Allosteric transitions
3) Ligand non covalent binding
4) Diffusion

Where equilibrium dynamics appears to be too short, advanced sampling methods exist: umbrella sampling, meta dynamics, etc.

Mechanics: main concepts

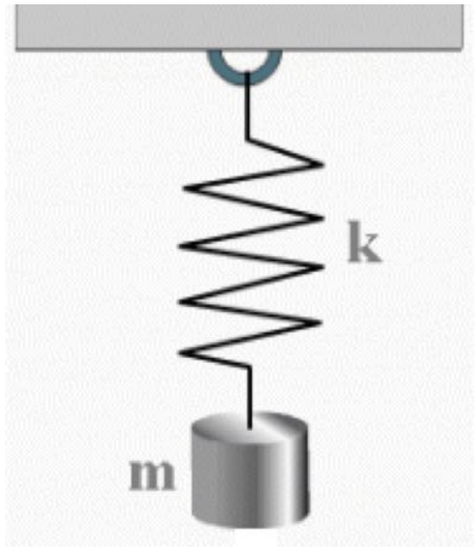
$$\vec{F} = m * \vec{a}$$ Newton's II law

$$\vec{a} = \frac{d\vec{v}}{dt} = \frac{d^2\vec{r}}{dt^2}$$ Differential form

$$\frac{d^2\vec{r}}{dt^2} = \vec{F}/m$$

$$\vec{F} = -grad(U)$$ Potential energy

$$H = T + U$$ Total energy, or Hamiltonian

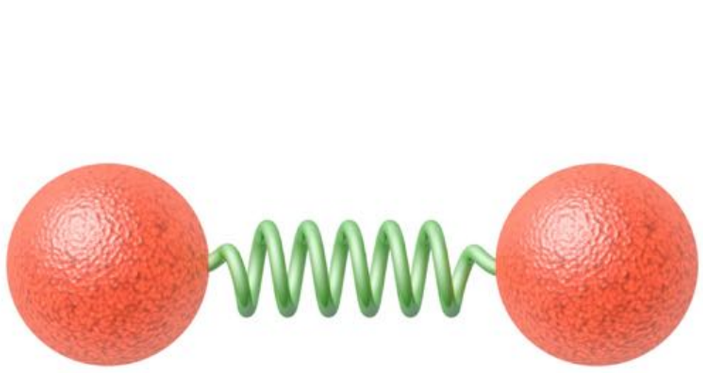


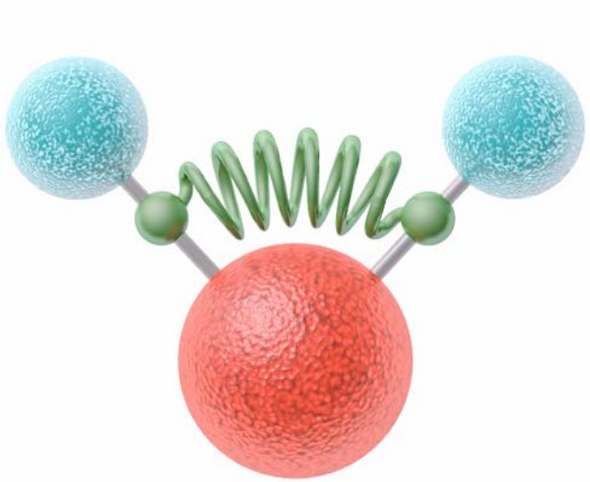
$$U = \frac{kx^2}{2}$$

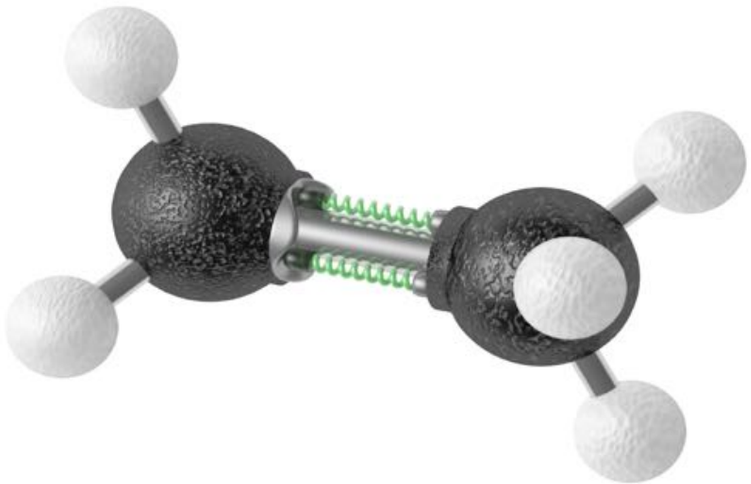
$$T = \frac{mv^2}{2}$$

Typical biomolecular force field, p. 1

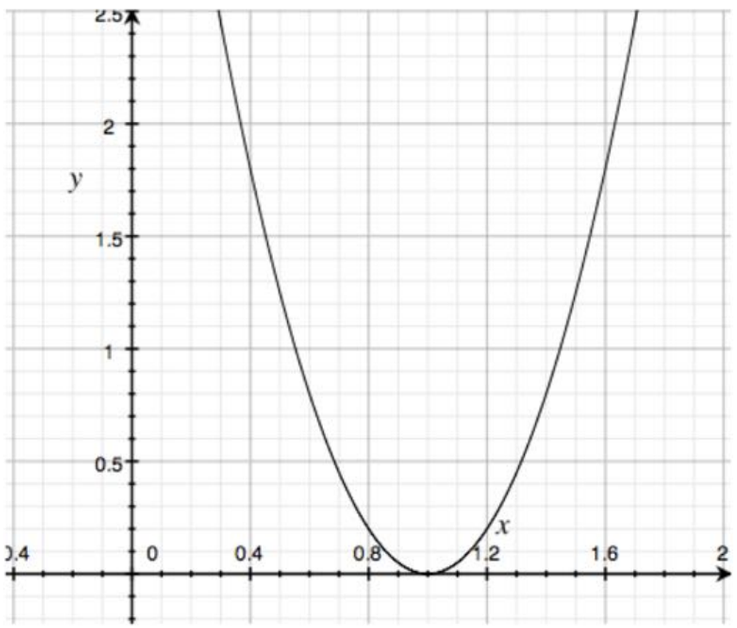
Bonded terms

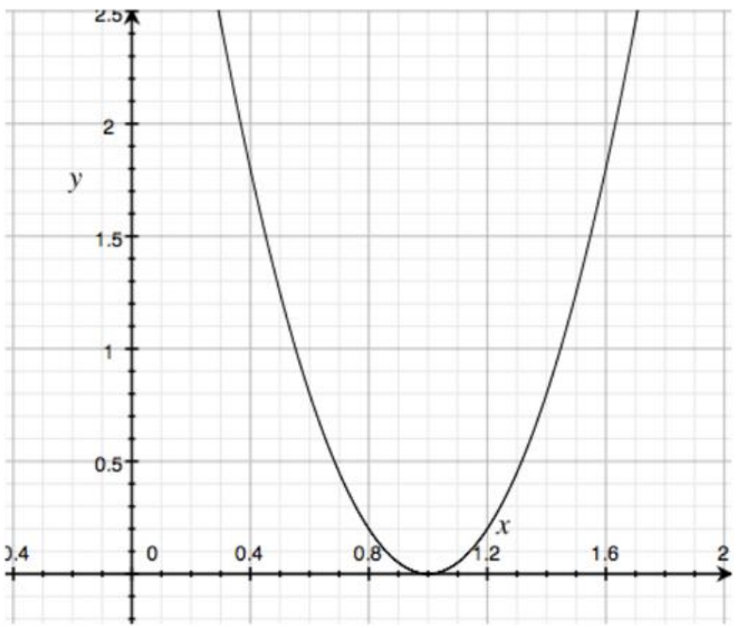


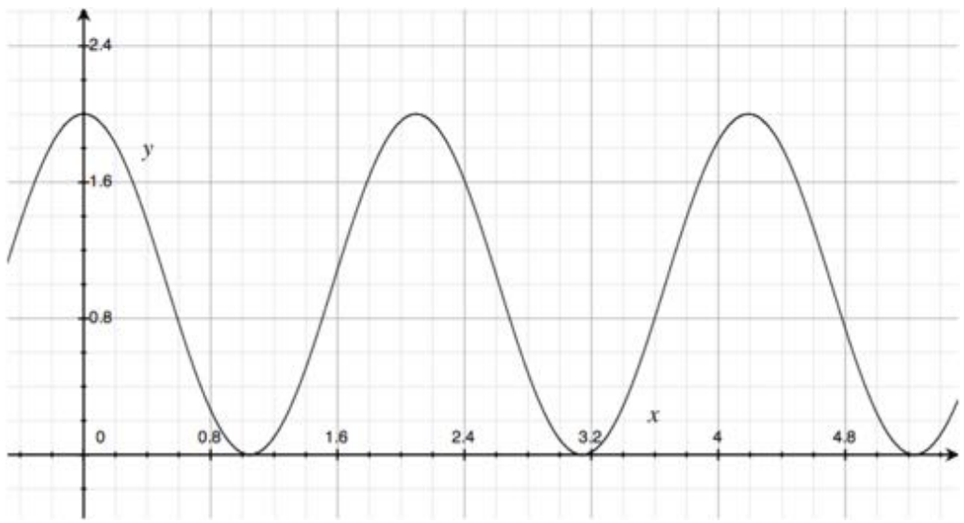




$$U_{bond} = \frac{k_b(l - l_0)^2}{2} \quad U_{ang} = \frac{k_a(\theta - \theta_0)^2}{2} \quad U_{tor} = V[1 + cos(n\phi - \phi_0)]$$





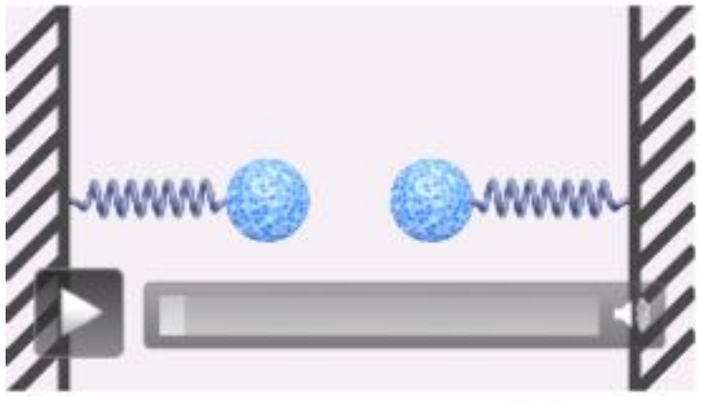


Typical biomolecular force field, p.2

Non-bonded terms

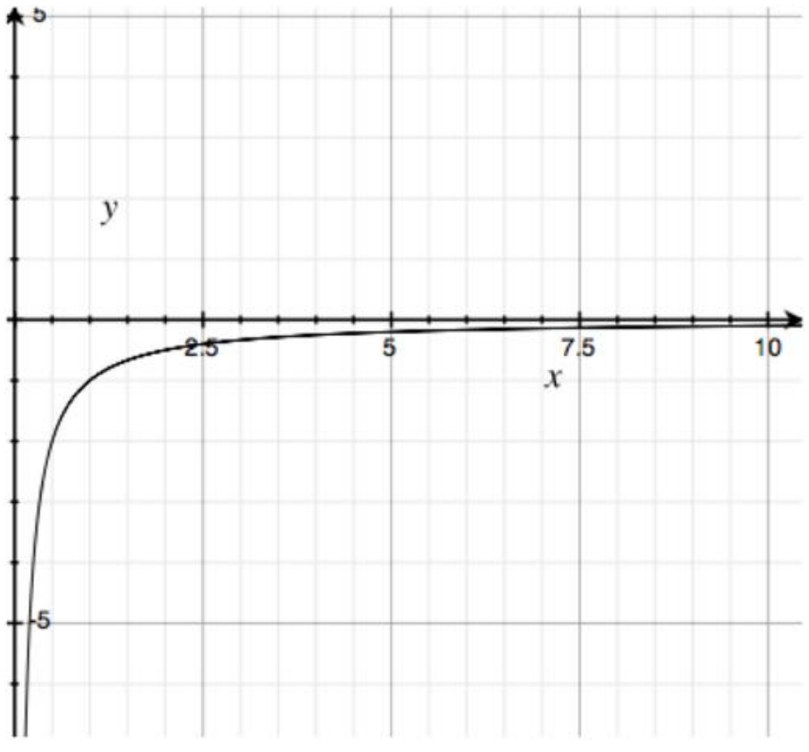
Electrostatic (Coulomb)

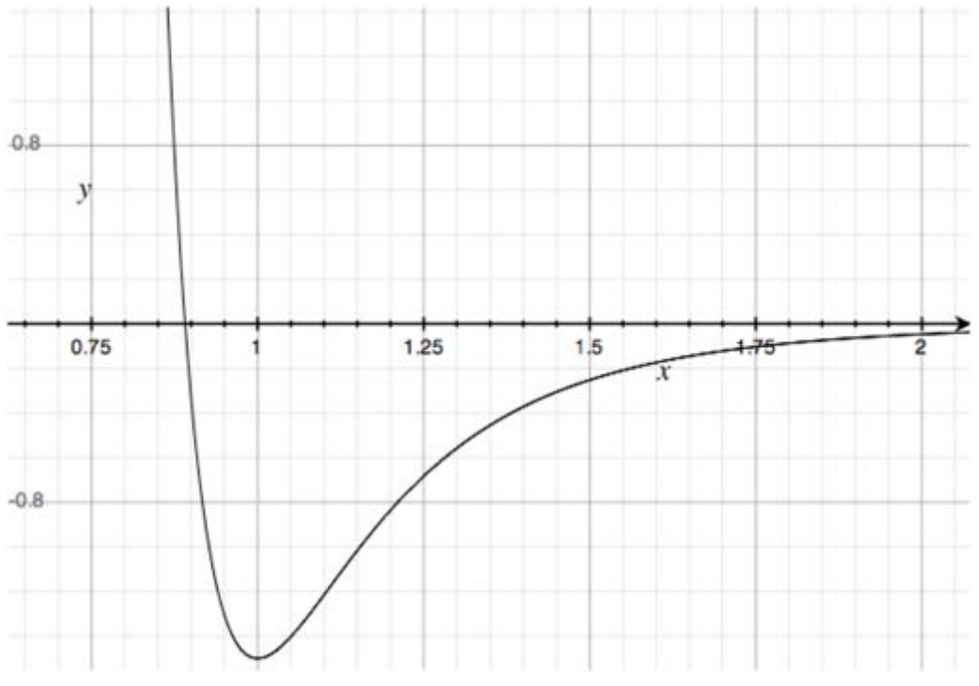
$$U_q = \frac{q_i q_j}{r}$$



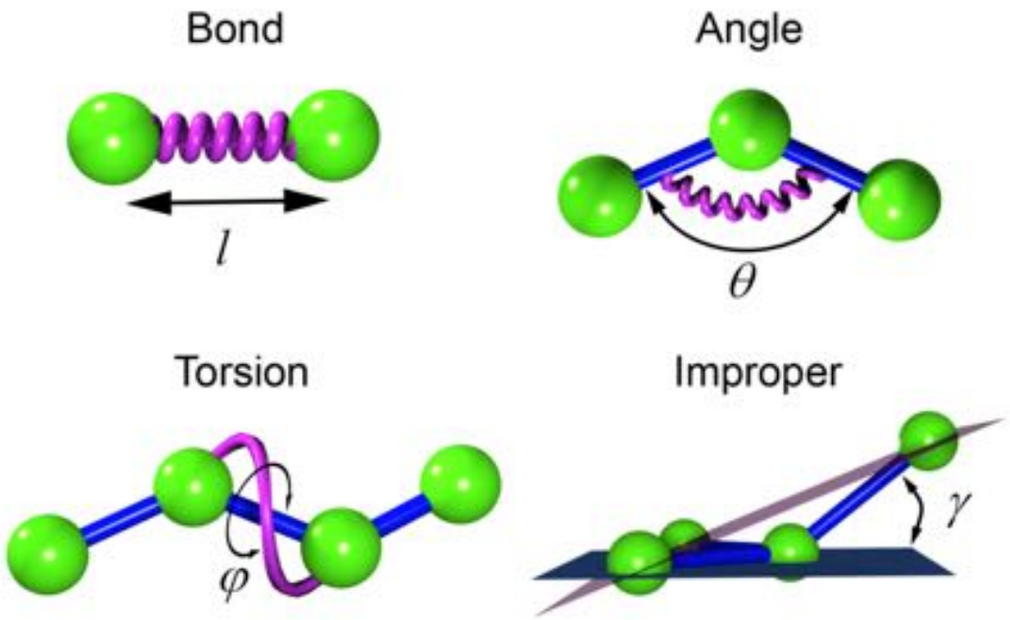
Lennard-Jones potential

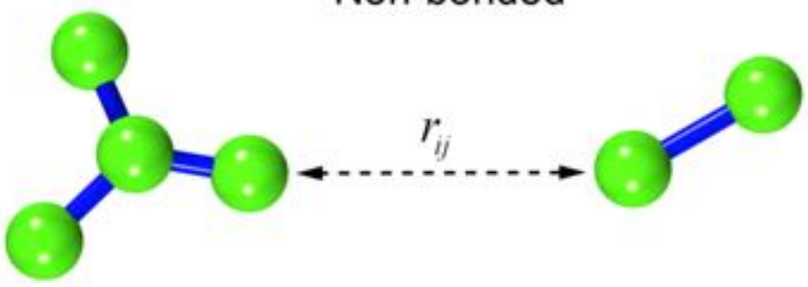
$$U_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right]$$





Typical biomolecular force field, p. 3





$$U(\{\vec{r}_i\}) = \sum_{bonds} \frac{1}{2} k_b (l - l_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n [1 + \cos(n\varphi - \varphi_0)]$$

$$+ \sum_{improper-cn} \frac{1}{2} k_\gamma (\gamma - \gamma_0)^2 + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} f_{ij}$$

Common force fields

For biomolecules:
CHARMM, AMBER, GROMOS, OPLS
Satellite general force fields (for ligands):
GAFF (for amber), GenFF (for charmm)

Corresponding water models:
TIP3P, SPC, TIP4P, TIP5P, SPCE/e, TIP4P-Ew, etc.

Many other force fields exist for other/ special purposes: e.g. CLASS 2 force fields (PCFF,COMPASS), polarizable force fields, reactive force fields (ReaxFF), coarse grain (MARTINI), united atom variations of AMBER and CHARMM, etc.

When choosing a force field it is better to consult original references!

How to get atoms moving?

$$\frac{d^2\vec{r}}{dt^2} = \vec{F}/m$$ Newton's II law is differential equation

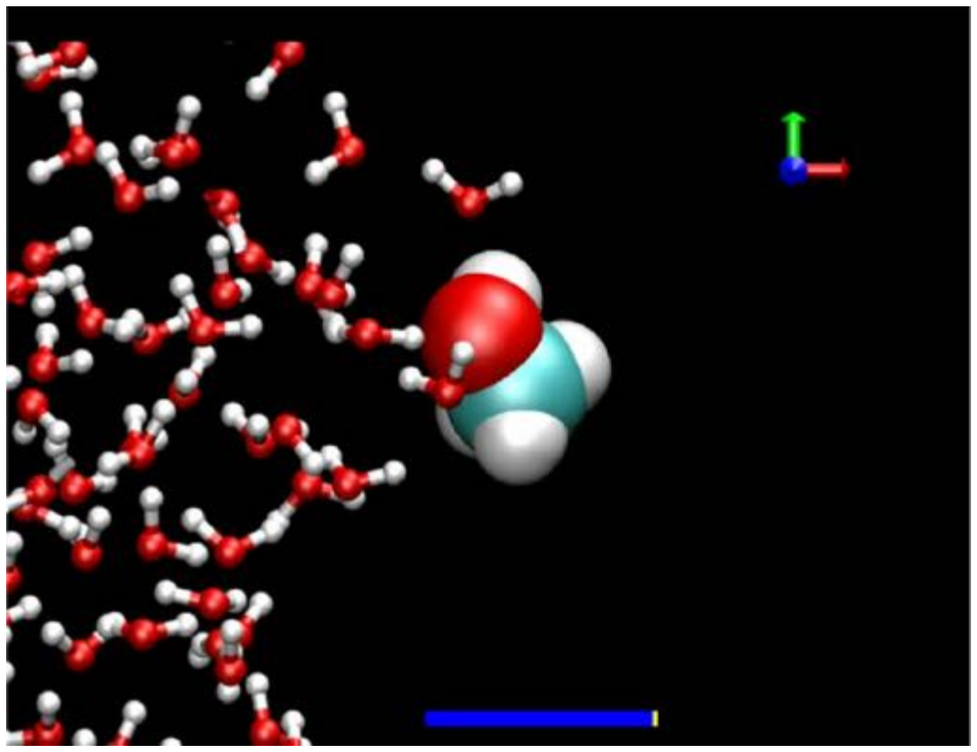
Solve it numerically!
Integrate it using a discrete time step!

$$\begin{cases} \frac{d\vec{r_i}}{dt} = \vec{v}_i \\ \frac{d\vec{v_i}}{dt} = \vec{F}_i/m_i \end{cases} \Longrightarrow \begin{cases} \frac{\vec{r_i}(t+\Delta t)-\vec{r_i}(t)}{\Delta t} = \vec{v}_i \\ \frac{\vec{v_i}(t+\Delta t)-\vec{v_i}(t)}{\Delta t} = \vec{F}_i(t)/m_i \end{cases}$$

Velocity Verlet integration scheme

(leap frog algorithm)

$$\begin{cases} \vec{r_i}(t+\Delta t) = \vec{r_i}(t) + \vec{v_i}(t)\,\Delta t + \frac{1}{2}\vec{a_i}\Delta t^2 \\ \vec{v_i}(t+\Delta t) = \vec{v_i}(t) + \frac{\vec{a_i}(t)+\vec{a_i}(t+\Delta t)}{2}\Delta t \end{cases}$$

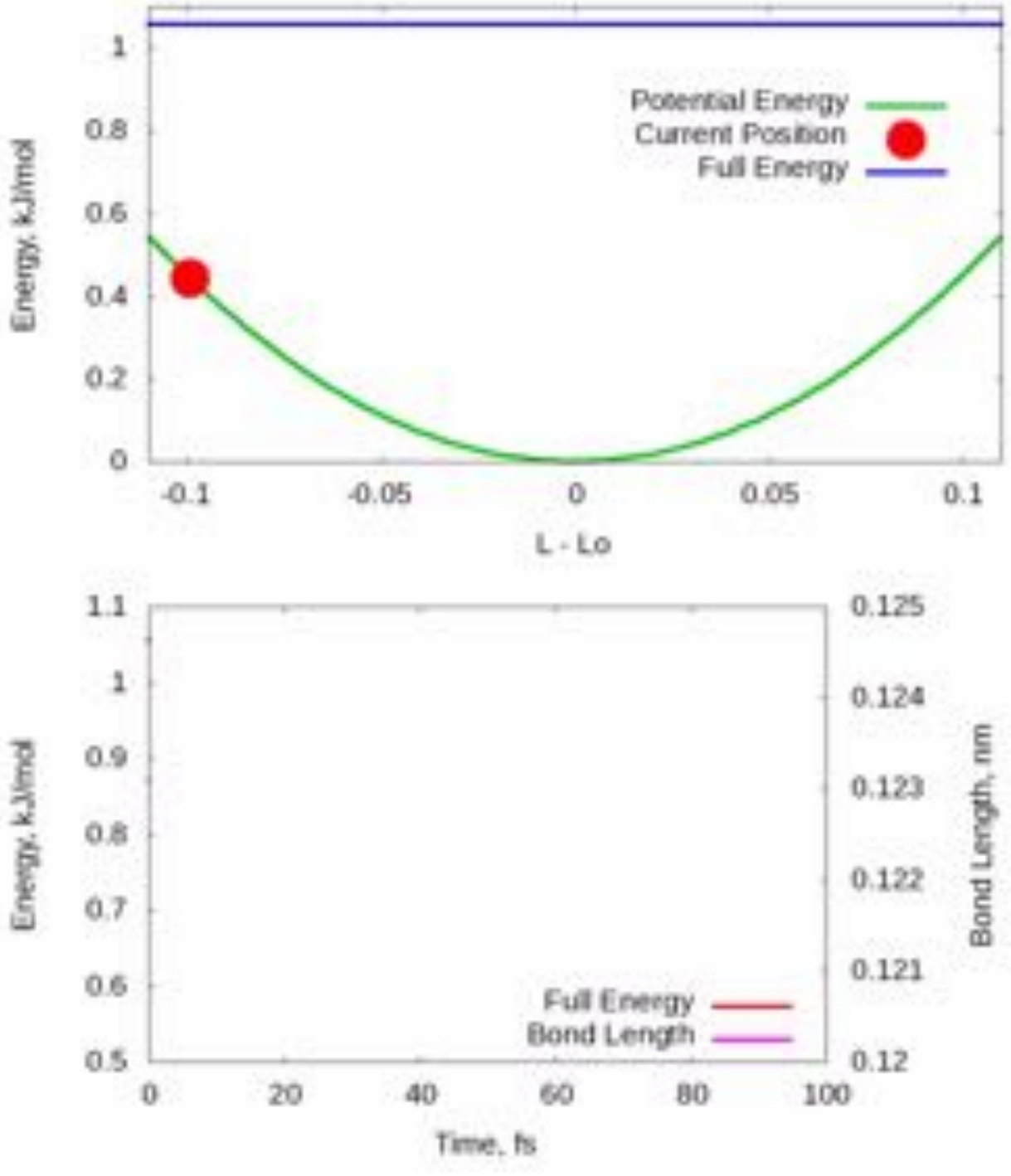


Example of MD simulation

- No long term energy drift (except for rounding errors)
- Nice stability
- Time reversible
- Global error of order O(t^2)

Close look at simple MD simulation

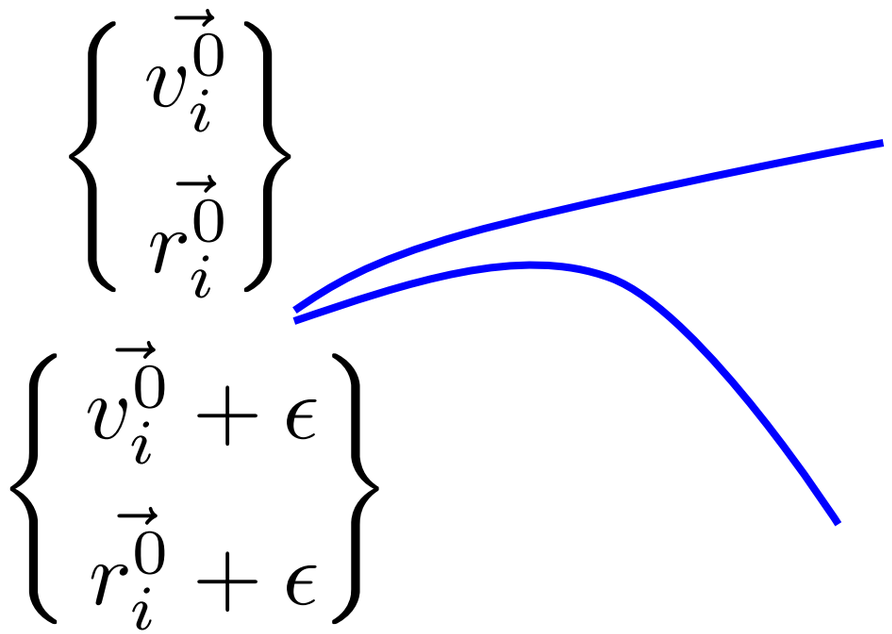
Oxygen molecule
NVE simulation
Leap-frog integrator
Integration step 1 fs



Is MD trajectory close to real one?

On short time scale – probably.
On long time scale – fundamentally not possible!

Lyapunov instability
of particle trajectories



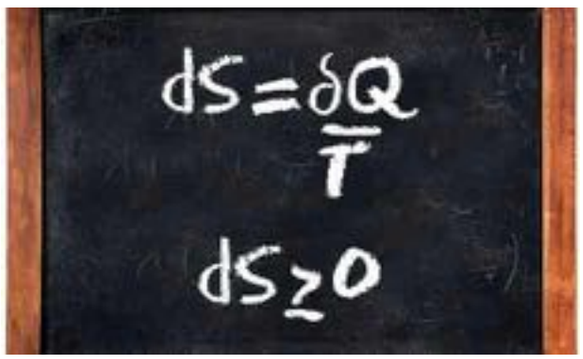
The system is chaotic!

Infinitesimal differences in initial conditions may drive system to different states!

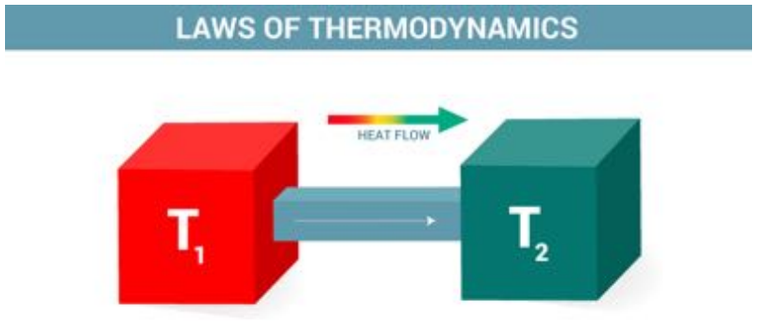
View MD as a statistical method – averages and evolution of averages are only meaningful on long time scale.

«Философское» отступление 1

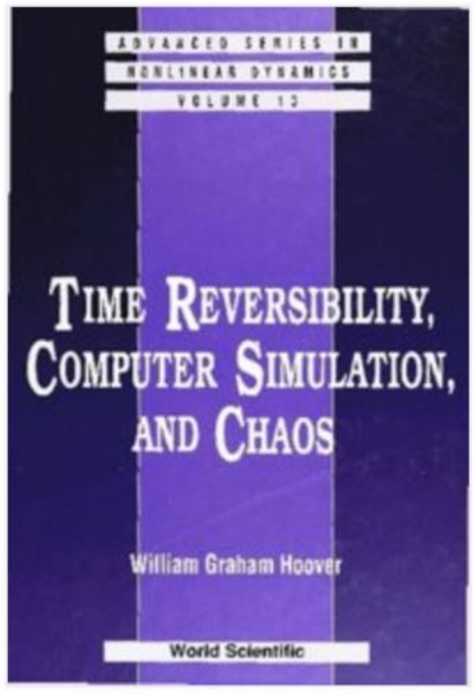
2 закон термодинамики vs 2 закон Ньютона



$$\vec{F} = m * \vec{a}$$



$$\vec{a} = \frac{d\vec{v}}{dt} = \frac{d^2\vec{r}}{dt^2}$$



«Философское» отступление 2

Демон Лапласа

> *Мы можем рассматривать настоящее состояние Вселенной как следствие его прошлого и причину его будущего. Разум, которому в каждый определённый момент времени были бы известны все силы, приводящие природу в движение, и положение всех тел, из которых она состоит, будь он также достаточно обширен, чтобы подвергнуть эти данные анализу, смог бы объять единым законом движение величайших тел Вселенной и мельчайшего атома; для такого разума ничего не было бы неясного и будущее существовало бы в его глазах точно так же, как прошлое.*

В философии с давних пор ведётся спор о существовании свободы воли, её определении и природе. Существуют две противоположные позиции:

- *метафизический либертарианизм* — утверждение, что детерминизм неверен и, таким образом, свобода воли существует или по меньшей мере возможна,
- *жёсткий детерминизм* — утверждение о том, что детерминизм верен и свободы воли не существует.

Пьер-Симон де Лаплас

Pierre-Simon de Laplace

Лаплас в мундире канцлера Сената. Фрагмент портрет кисти Жана-Батиста Герена, 1838

Дата рождения	23 марта 1749
Место рождения	Бомон-ан-Ож, Нормандия
Дата смерти	5 марта 1827 (77 лет)
Место смерти	Париж
Страна	Франция
Научная сфера	математика, механика, физика, астрономия

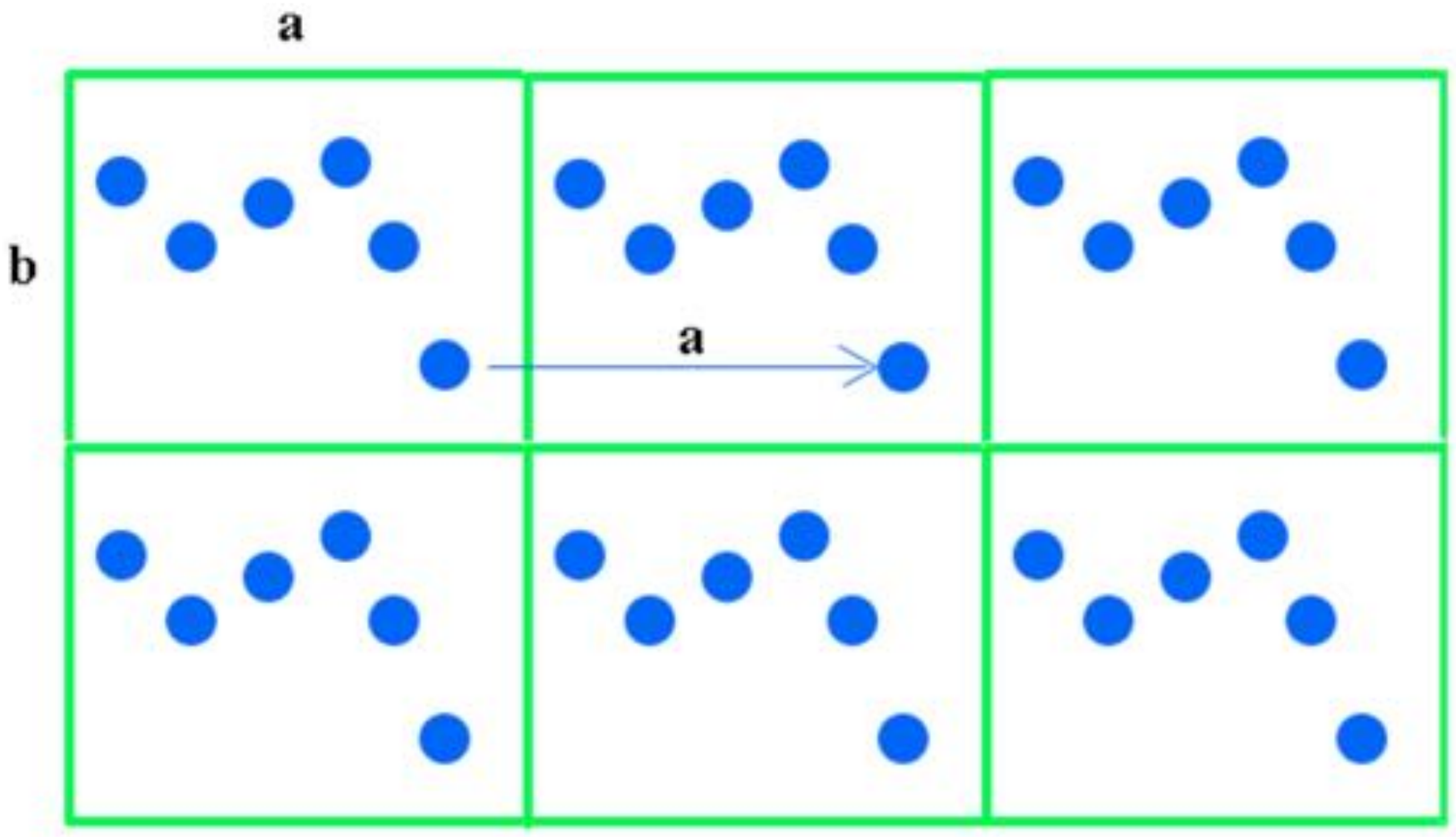
Common MD techniques

Needed almost in every simulation

- Periodic boundary conditions
- Interaction cut off
- Electrostatics
- Constant temperature/pressure simulations
- Rigid bonds
- Parallelization and HPC computing

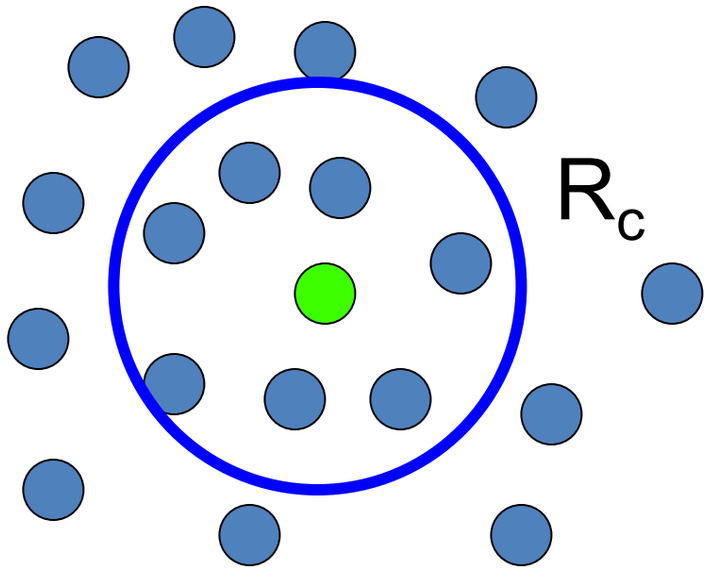
Periodic boundary conditions

A trade off to alleviate boundary effects



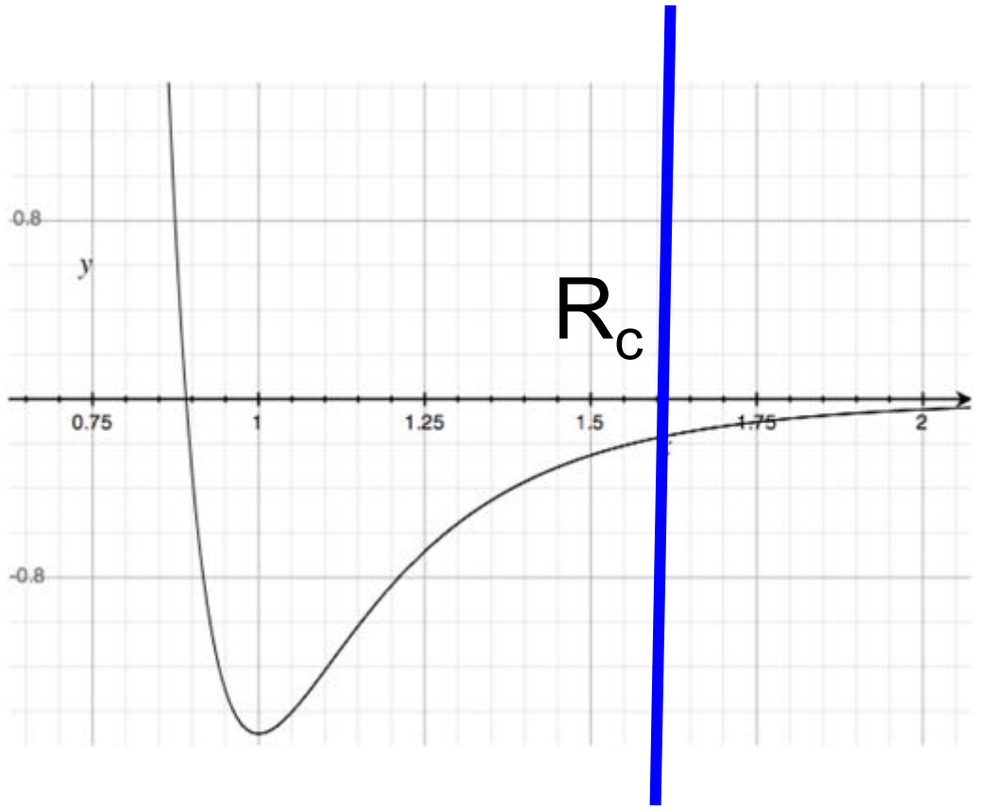
$$\vec{r}_{klm} = \vec{r} + k * \vec{a} + l * \vec{b} + m * \vec{c}$$

Non bonded interaction cut off



Simple cut off radius

Switching function may be used to zero the truncated potential smoothly over some additional distance beyond cut off



$$U_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right]$$

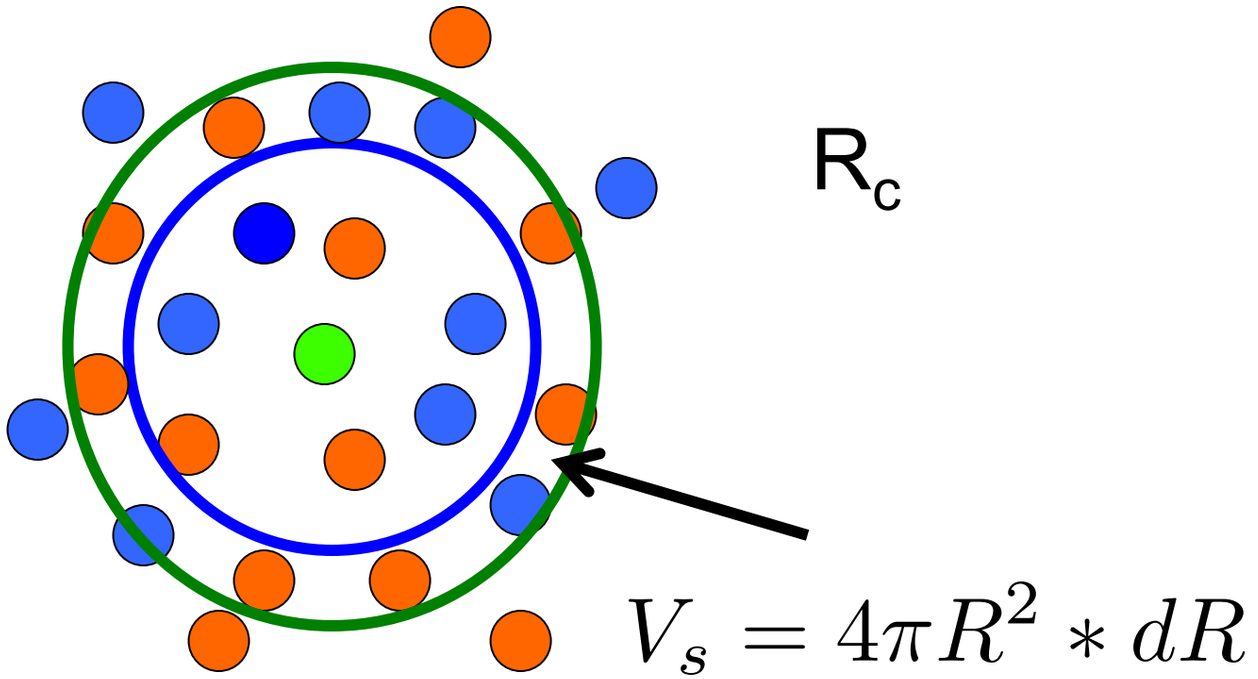
Lennard-Jones potential

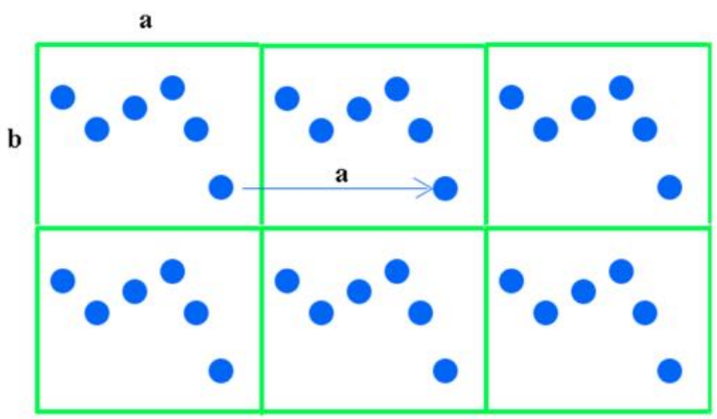
Is it fair to cut off electrostatics?

$$U_q = \frac{q_i q_j}{r}$$

Coulomb potential

$$\phi(r_i) = q\sum_{j+}^{\infty}\frac{1}{|r_i - r_{j+}|} - q\sum_{j-}^{\infty}\frac{1}{|r_i - r_{j-}|}$$

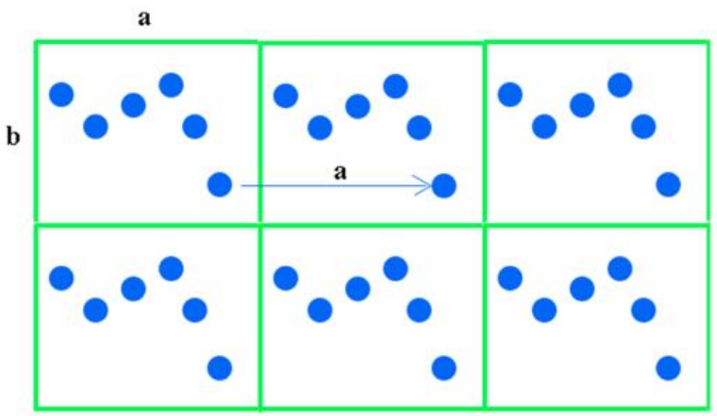




Using plain cut offs for Coulomb interactions is known to be bad for charged systems

Eg.: DNA will be unstable, lipid simulations will behave badly, etc.

Ewald summation method

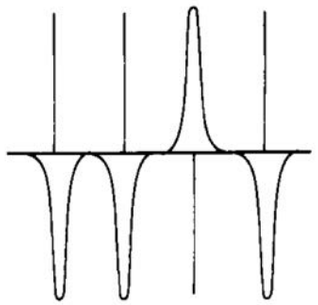


$$U = \frac{1}{2}\sum_{i,j.n}\frac{q_i q_j}{\left|r_i - r_j + nL\right|}$$

How to calculate it?

$$-\nabla^2\phi(\mathbf{r}) = 4\pi\rho(\mathbf{r})$$

Poisson equation – can be solved for periodic systems in Fourier space



Add and subtract a Gaussian smeared charge

$$\rho_{Gauss}(r) = -q_i(\alpha/\pi)^{3/2}\exp(-\alpha r^2)$$

Real part

Fourier part

$$U = \frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j \operatorname{erfc}(\sqrt{\alpha} r_{ij})}{r_{ij}} + \frac{1}{2V}\sum_{\mathbf{k}\neq 0}\frac{4\pi}{k^2}\left|\rho(\mathbf{k})\right|^2\exp(-k^2/4\alpha) - (\alpha/\pi)^{1/2}\sum_{i=1}^{N}q_i^2$$

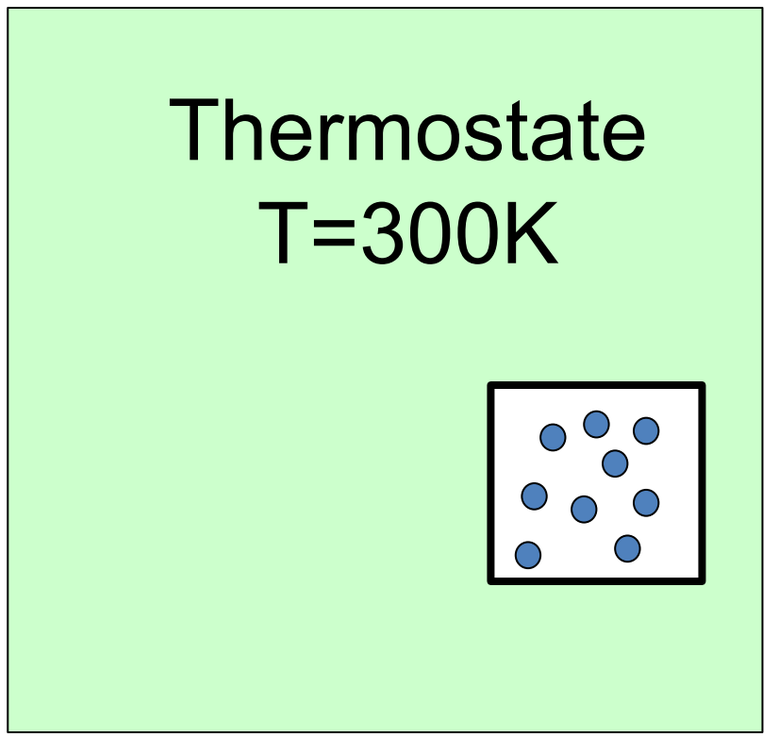
Faster modification: Particle Mesh Ewald (**PME**) method, makes use of grid interpolation and fast Fourier transforms, scales as N*log(N). This is an accepted standard for MD simulations now.

Constant temperature simulations

- Why use temperature coupling?
- -Compensate energy drift
- -NVT ensemble

$$P_{microstate} \propto e^{-\frac{H}{kT}}$$

Gibbs distribution



Two possibilities:
1)Add stochastic interactions
2)Use modified equations of motion

Constant temperature simulations 2

Stochastic approach

Andersen thermostat
-random collision with virtual particles

$$P(t) = \nu e^{-\nu t}$$

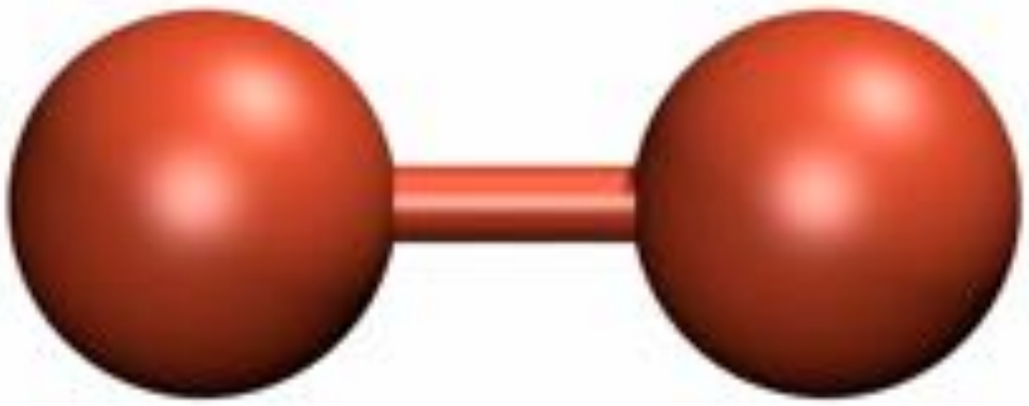
Langevin stochastic dynamics
-add friction and stochastic term

$$m\frac{d^2\vec{r_i}}{dt^2} = \vec{F_i} - \lambda\frac{d\vec{r_i}}{dt} + \boldsymbol{\eta}\left(t\right)$$

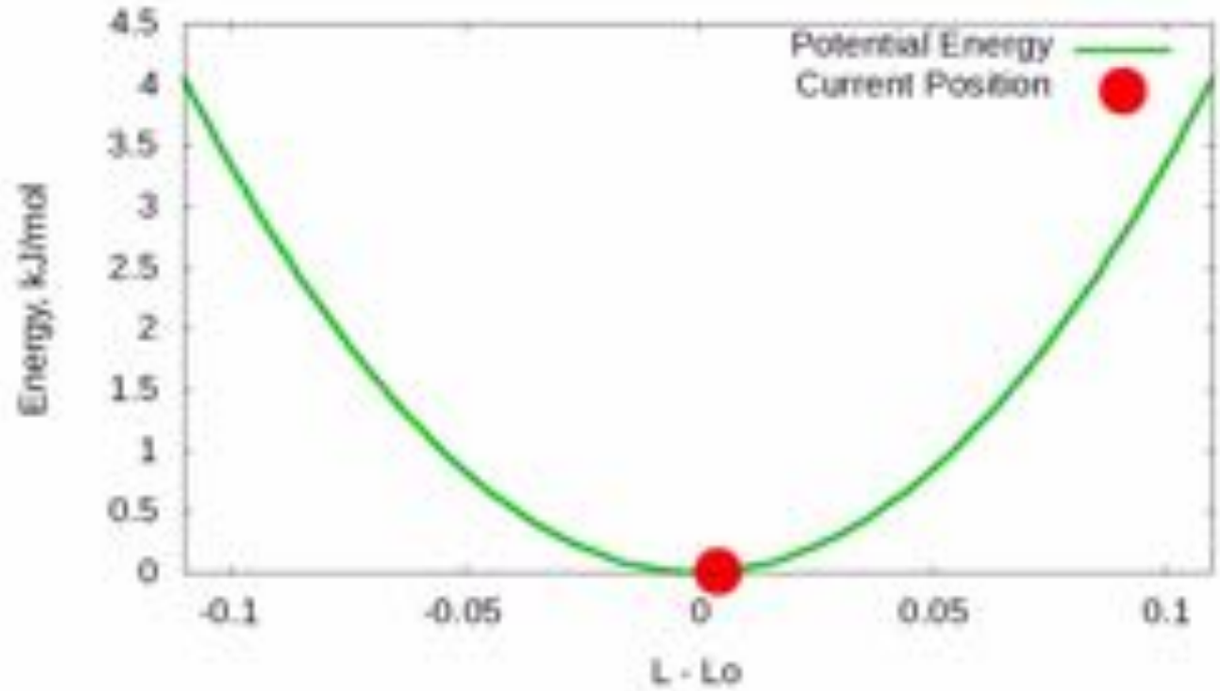
$$\left\langle \eta_i\left(t\right)\eta_j\left(t'\right)\right\rangle = 2\lambda k_B T\delta_{i,j}\delta\left(t-t'\right)$$

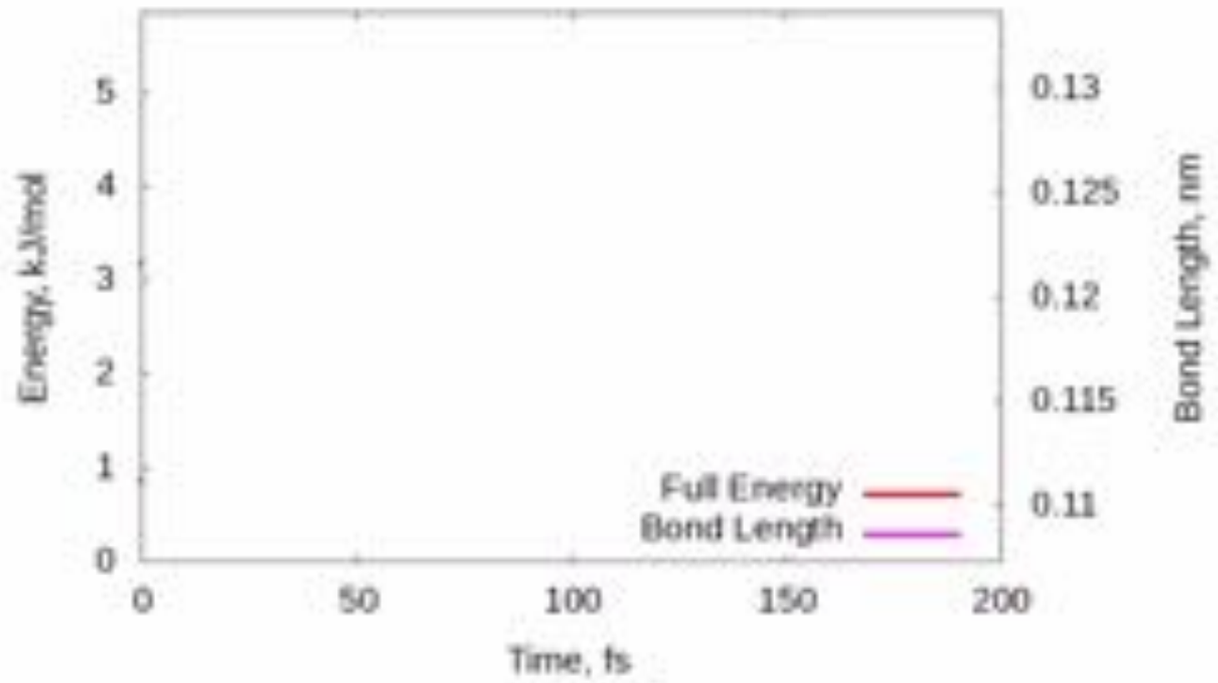
Constant temperature simulations 2

Langevin dynamics



Oxygen molecule
Stochastic dynamics simulation





Constant temperature simulations 3

Berendsen thermostat

$$\frac{kT}{2}(3N - N_c) = \sum_{i=1}^{N} m_i \vec{v_i}^2$$

-instantaneous temperature

Modify equations of motion

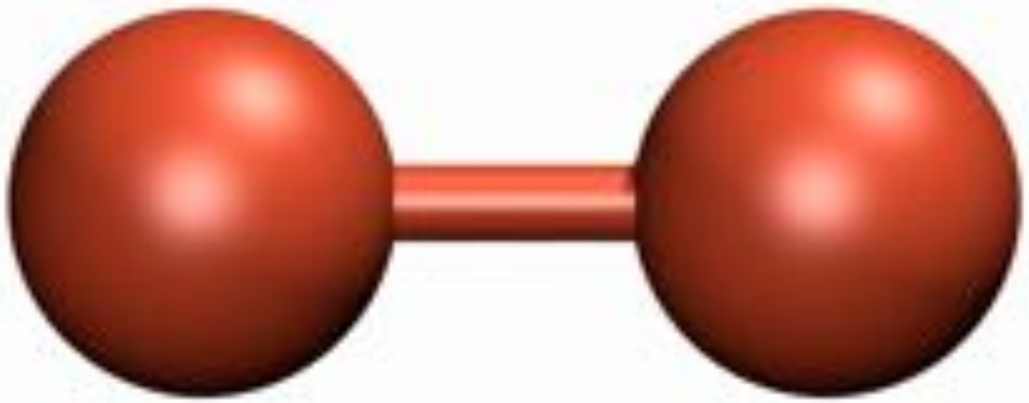
-re-scale the velocities of particles, multiply by

$$\lambda = \sqrt{T_0/T(t)}$$

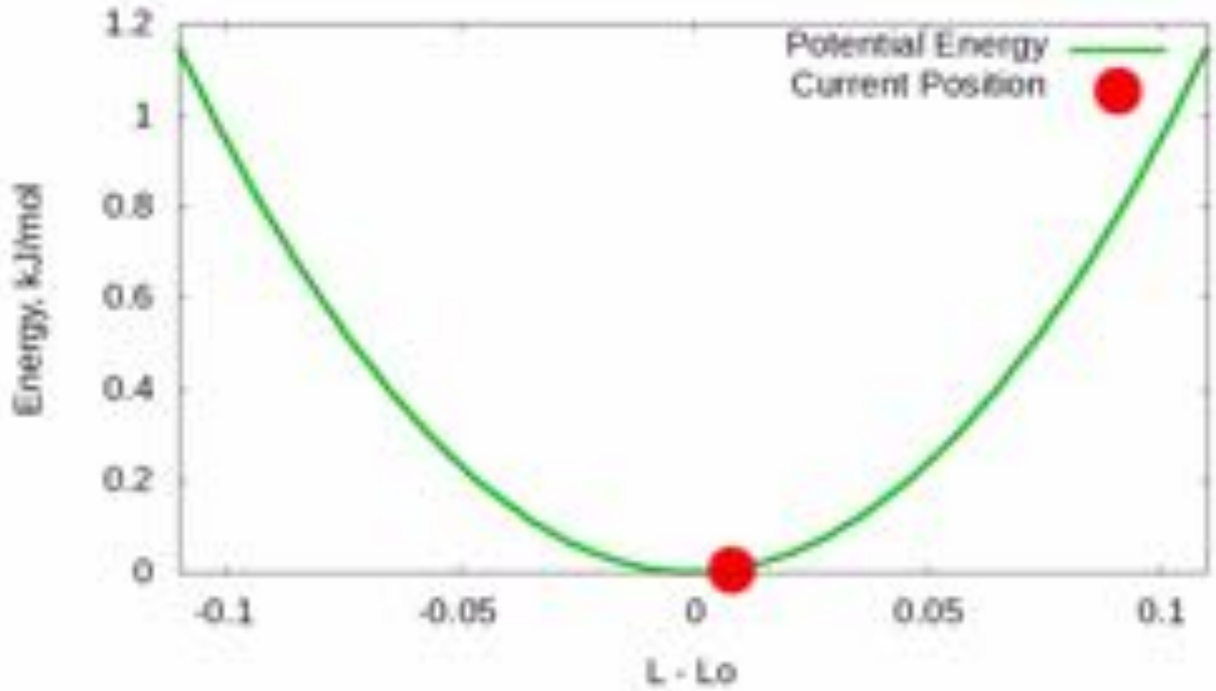
(!) incorrect ensemble, flying ice cube effect may happen

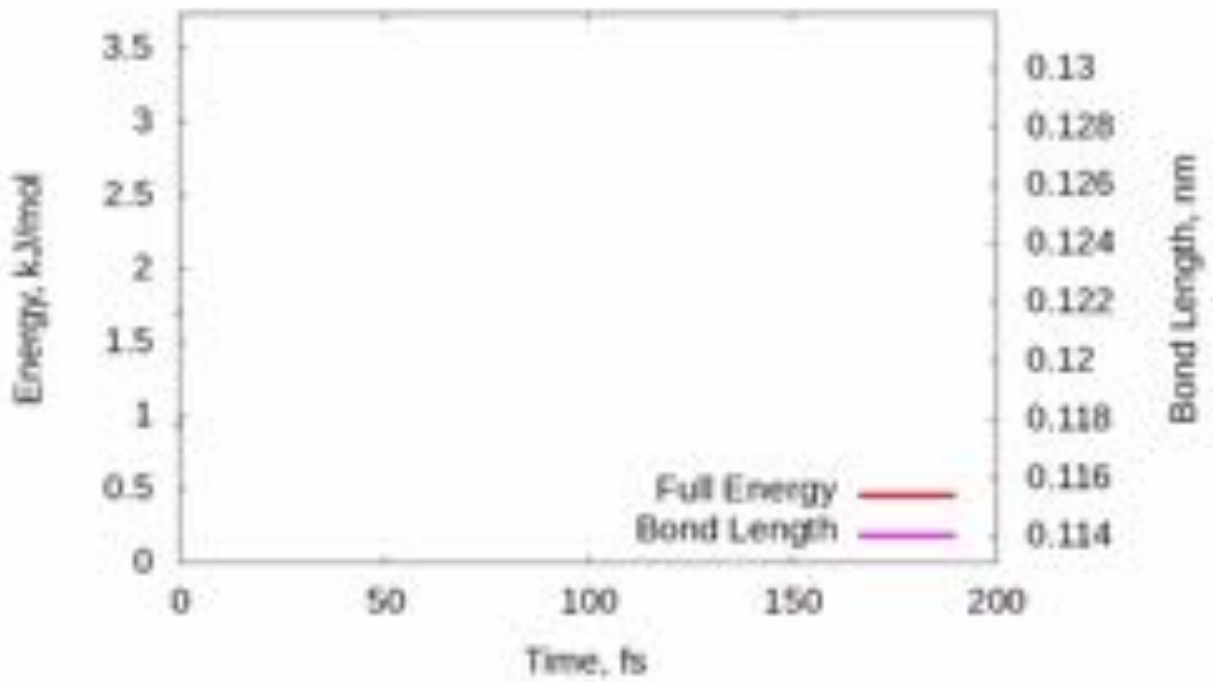
Constant temperature simulations 3

Berendsen thermostat



Oxygen molecule
NVT simulation
Leap-Frog integrator
Integration step 1 fs
Berendsen thermostat





Constant temperature simulations 3

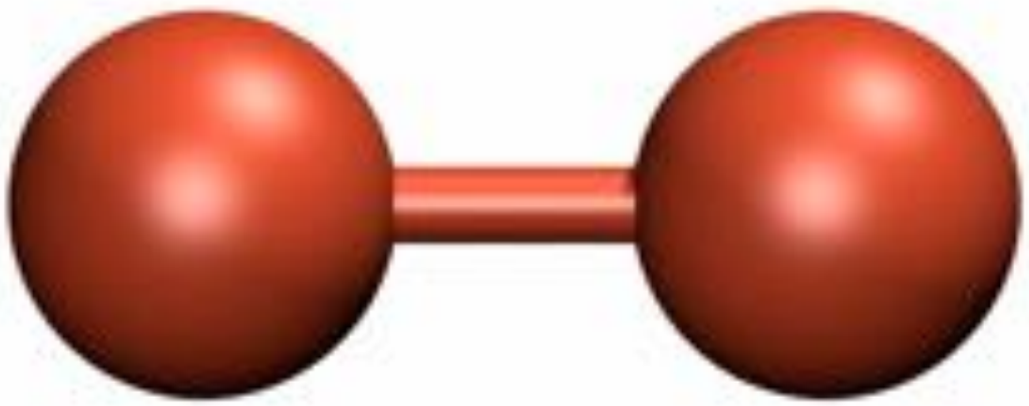
Nose-Hoover thermostat

Extended system approach
-add a “bath particle”
-modify equations of motion, Non-Hamiltonian dynamics
-gives NVT ensemble if:
1) No external forces, 2) there is only one conservation law, 3) the system is ergodic

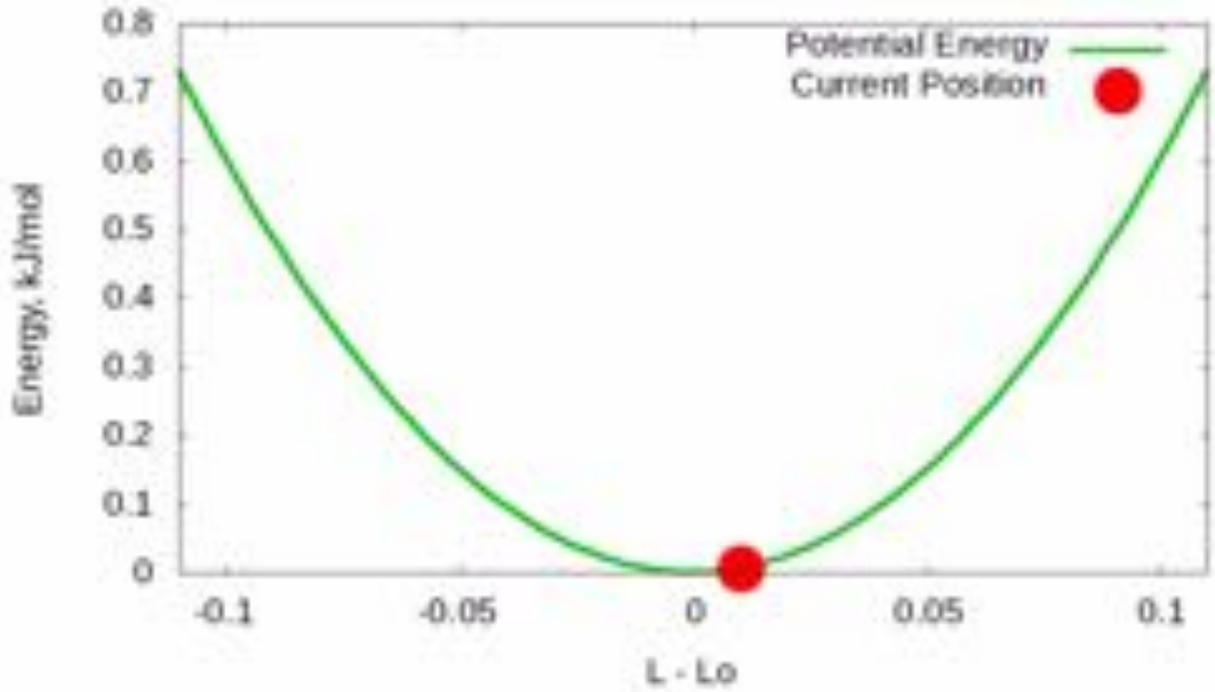
$$\begin{cases} \dfrac{d^2\vec{r_i}}{dt^2} = \dfrac{F_i}{m_i} - \dfrac{d\xi}{dt}\dfrac{d\vec{r_i}}{dt} \\ \dfrac{d^2\xi}{dt^2} = \dfrac{N_f k}{Q}\left[T - T_0\right] \end{cases} \qquad E_{NH} = E_k + U + \frac{1}{2}Q\dot{\xi}^2 + N_f kT\xi$$

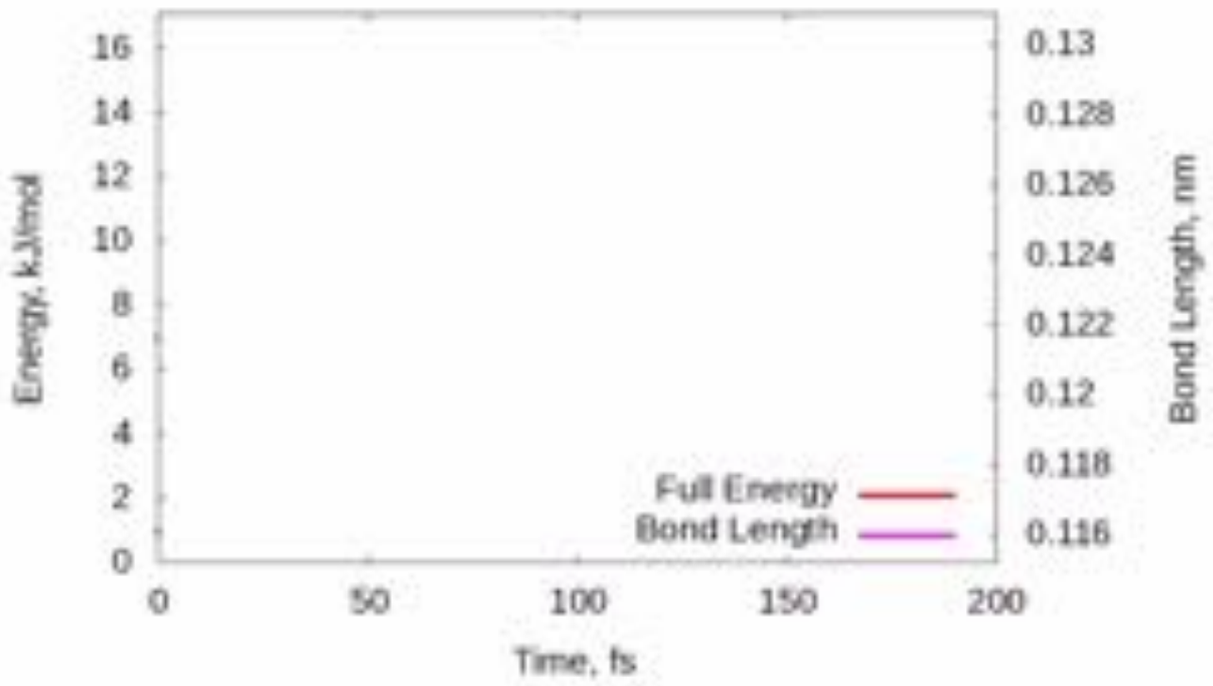
Constant temperature simulations 3

Nose-Hoover thermostat



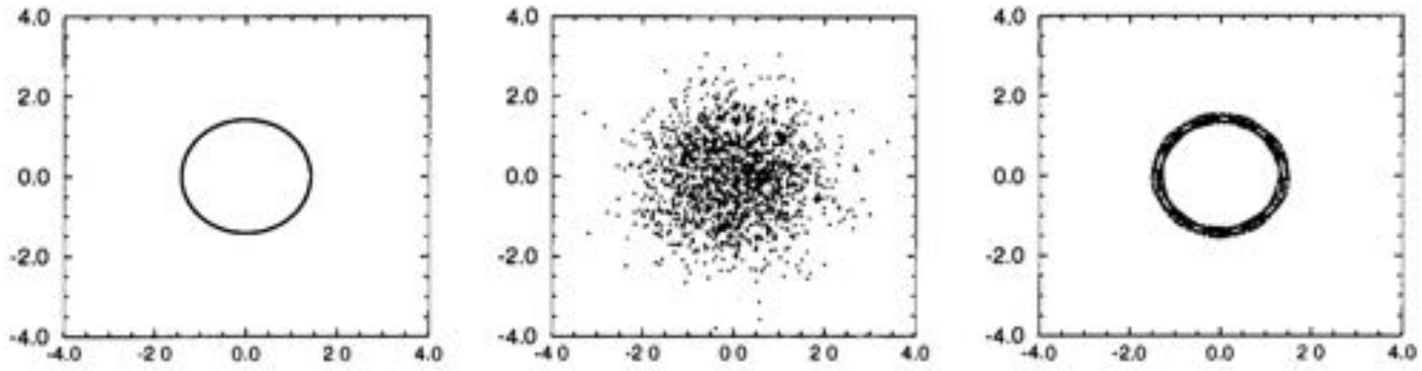
Oxygen molecule
NVT simulation
Verlet integrator
Integration step 1 fs
Nose-Hoover thermostat





Constant temperature simulations 3

Nose-Hoover chains



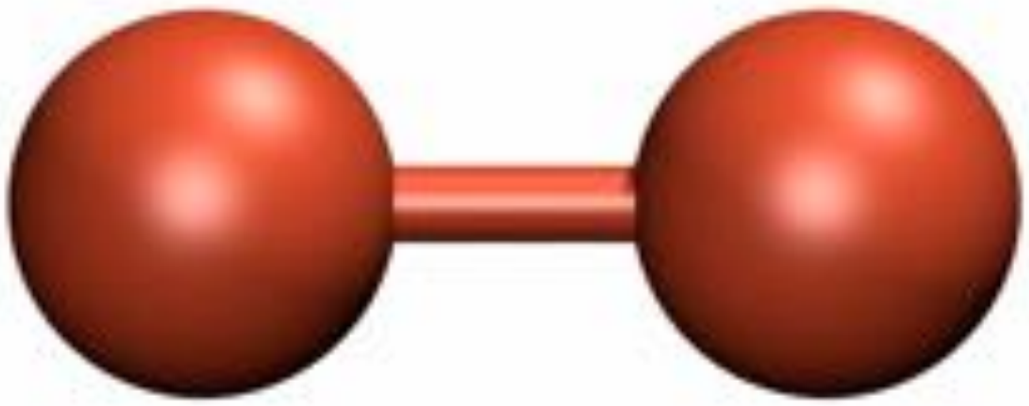
From D. Frenkel, B. Smit “Understanding Molecular Simulation, Second Edition: From Algorithms to Applications”

Nose-Hoover will not be correct for an oscillator, and “harmonic” systems.
Use NH chains instead!

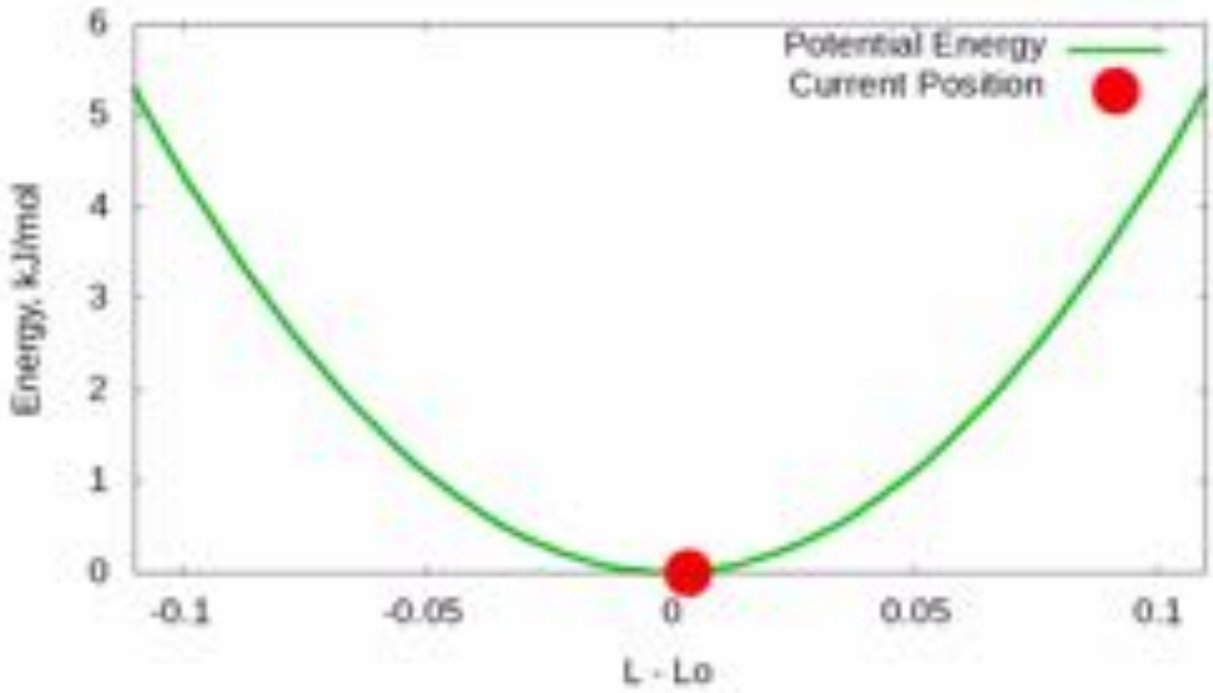
$$
\begin{aligned}
\dot{\boldsymbol{r}}_i &= \frac{\boldsymbol{p}_i}{m_i} \\
\dot{\boldsymbol{p}}_i &= \boldsymbol{F}_i - \boldsymbol{p}_i\dot{\xi}_1 \\
\ddot{\xi}_1 &= \frac{1}{Q_1}\left[\sum_{i=1}^{N} m_i \boldsymbol{v}_i^2 - N_f k_B T_{set}\right] - \dot{\xi}_1\dot{\xi}_2 \\
\ddot{\xi}_j &= \frac{1}{Q_j}\left[Q_{j-1}\dot{\xi}_{j-1}^2 - k_B T_{set}\right] - \dot{\xi}_j\dot{\xi}_{j+1} \quad , j = 2, \ldots, M-1 \\
\ddot{\xi}_M &= \frac{1}{Q_M}\left[Q_{M-1}\dot{\xi}_{M-1}^2 - k_B T_{set}\right] \quad ,
\end{aligned}
$$

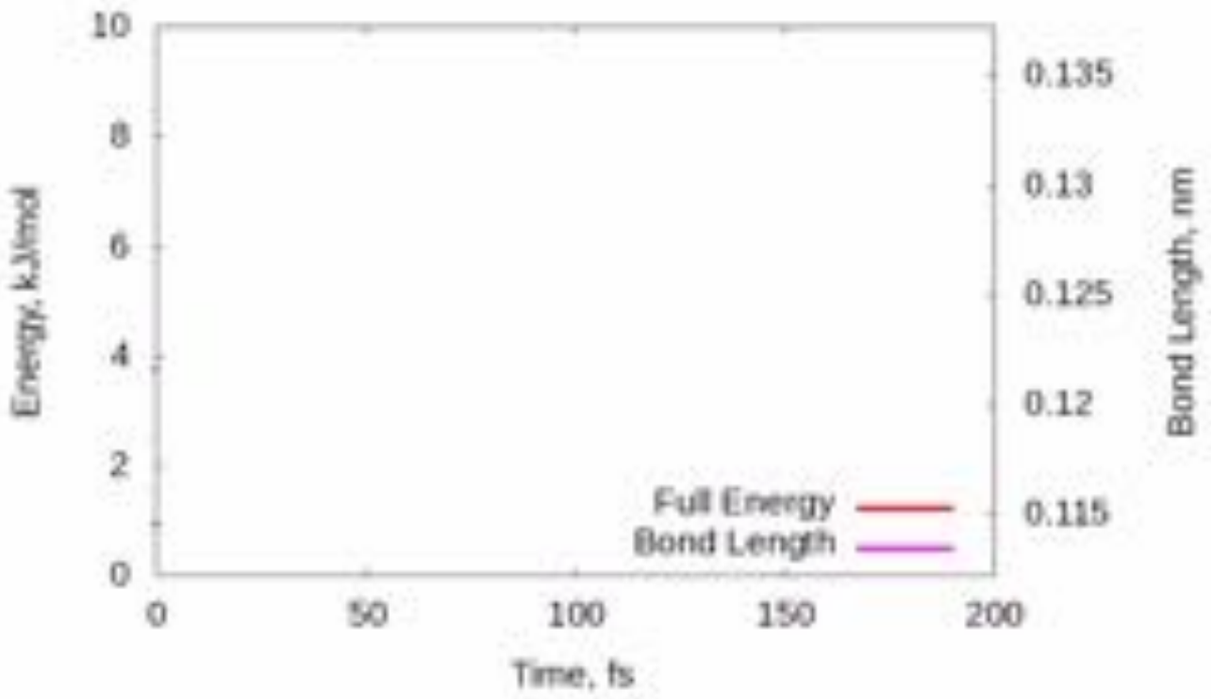
Constant temperature simulations 3

Nose-Hoover chains



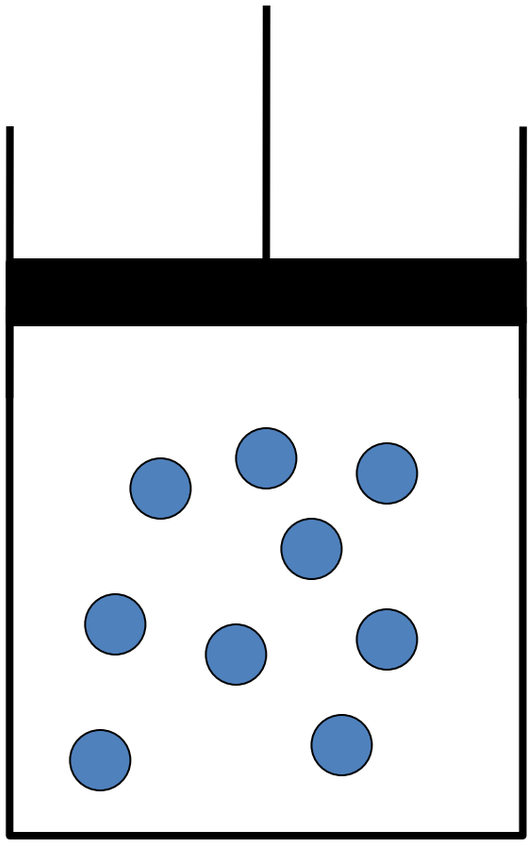
Oxygen molecule
NVT simulation
Verlet integrator
Integration step 1 fs
Nose-Hoover + chains thermostat





Constant pressure simulations

Berendsen Barostat



NPT-ensemble

$$P = nkT + vir/V$$

$$vir = \frac{1}{3}\sum \vec{f_{ij}}\vec{r_{ij}}$$

$$\begin{cases} \vec{r_i} \Longrightarrow \mu\vec{r_i} \\ L \Longrightarrow \mu L \end{cases}$$

$$\mu = \left[1 - \frac{\Delta t}{\tau_p}(P - P_0)\right]$$

A more robust and flexible algorithm is Parrinello-Rahman method, gives correct ensemble.

Rigid bonds

- Special algorithms (SHAKE,RATTLE,SETTLES, etc.) can keep bonds between the atoms fixed. (Holonomic constraints)
- This usually allows to increase integration timestep to 2 fs.

Computational costs of MD cycle

Initial coordinates

Force calculation

$$\boldsymbol{F}_i = -\frac{\partial V}{\partial \boldsymbol{r}_i} \qquad \boldsymbol{F}_i = \sum_j \boldsymbol{F}_{ij}$$

Coordinate update

$$\frac{\mathrm{d}^2 \boldsymbol{r}_i}{\mathrm{d}t^2} = \frac{\boldsymbol{F}_i}{m_i}$$

Data output

Computational time

$$\sum_{\theta} K_\theta(\theta - \theta_0)^2 +$$
$$\sum_{\phi} K_\phi(1 + cos(n\phi - \phi_0))$$

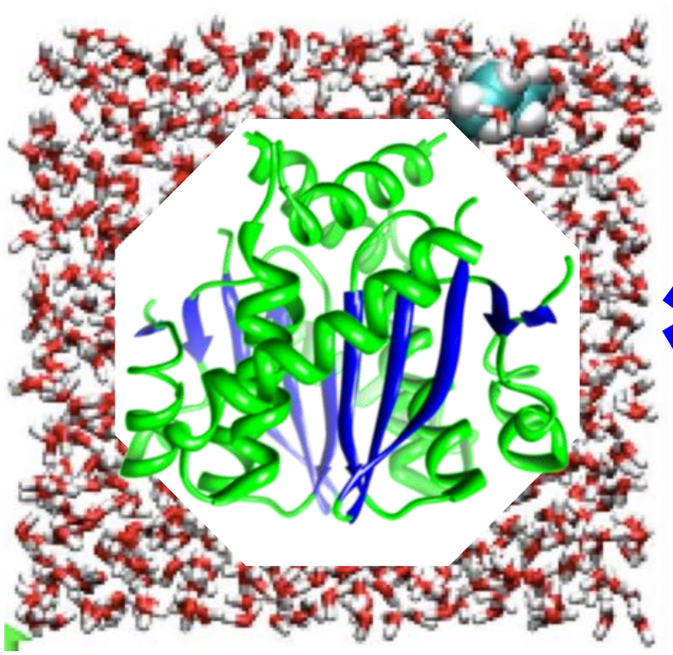
~2 %

$$\sum \epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] +$$
$$\sum_{ij} \frac{q_i q_j}{r_{ij}}$$

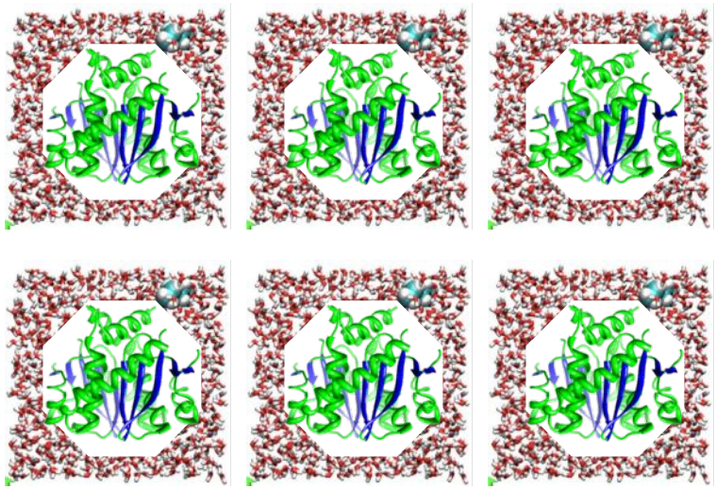
~90 %

~0.1 %

Parallel computing

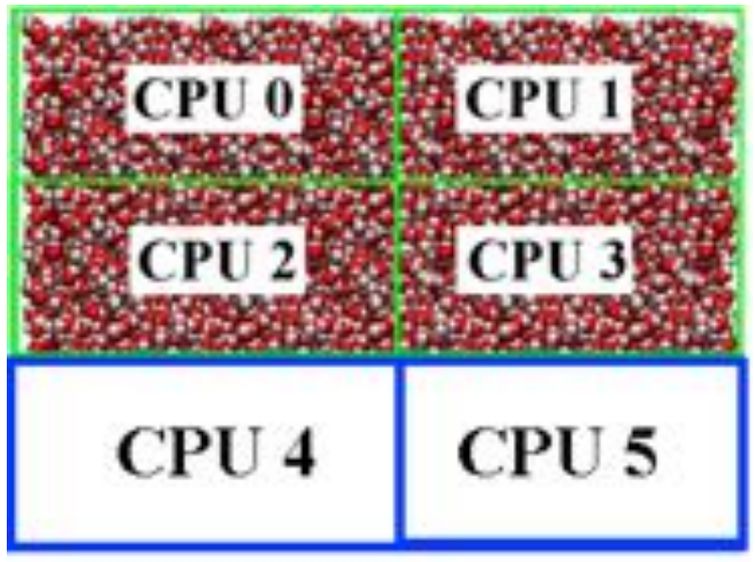


System under study



Multiple replica calculation

GRIDs, distributed computing, etc



Domain/task decomposition

supercomputing (general, with GPU, ASIC, etc)

Special architectures

Specially designed Application Specific Integrated Circuits (ASIC):

- MD GRAPE, RIKEN, Japan
- ANTON computer, D.E. Shaw Research, USA

David Shaw

Born	David Elliot Shaw March 29, 1951 (age 67)
Alma mater	University of California, San Diego Stanford University
Occupation	Investor, computer scientist, and hedge fund manager
Known for	Founding and managing D. E. Shaw & Co.
Net worth	US$ 6.2 billion (June 2018)[1]

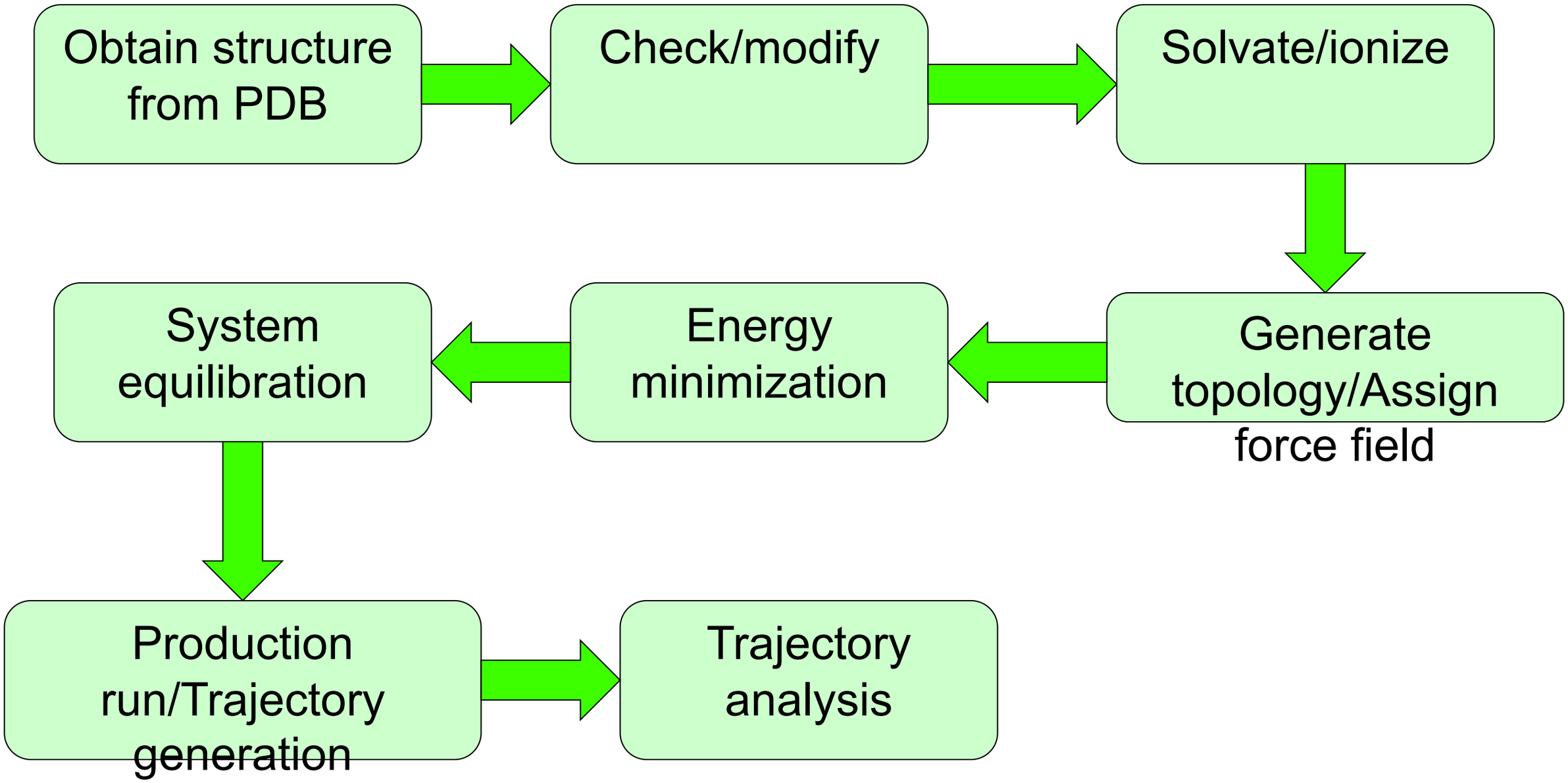
Chemical system (PDB ID)	Number of atoms	Approximate performance (microseconds/machine-day)*
DHFR (5DFR)	23558	17.4
aSFP (1SFP)	48423	11.7
FtsZ (1FSZ)	98236	5.7
T7Lig (1A01)	116650	5.5

http://www.psc.edu/index.php/computing-resources/anton

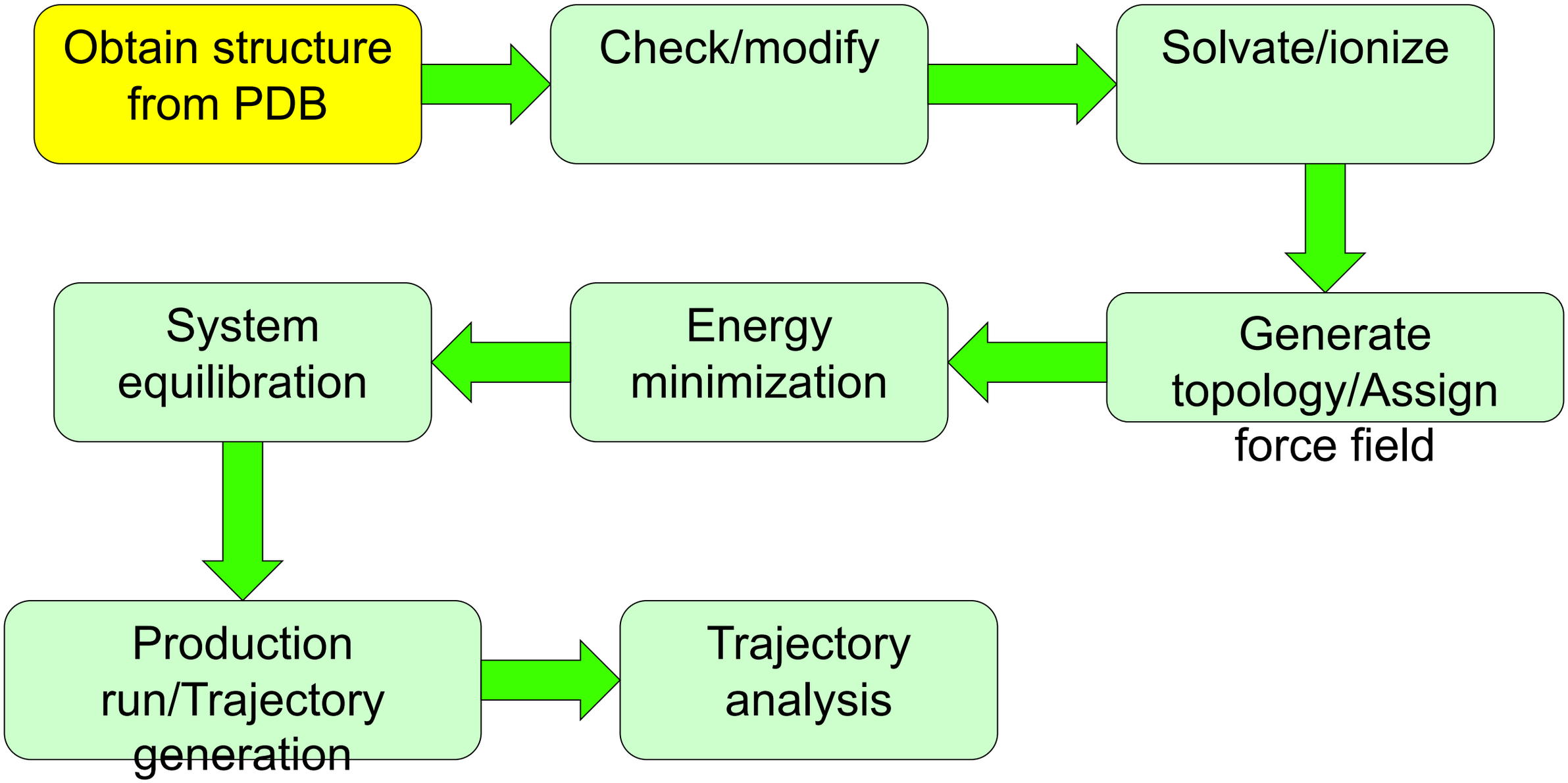
Software for MD simulations

- CHARMM
- AMBER
- NAMD
- GROMACS
- DESMOND
- TINKER
- LAMMPS
- DL_POLLY
- GROMOS
- CPMD, CP2K, …

How do I do an MD simulation?



PDB and MM crystallography



PDB and MM crystallography

- Google PDB – go to PDB site
- Type 1ubq , press search
- Press 3D-View
- select Structure -> “asymmetric unit”; Crystal packing -> “3x3 unit cell …”
- Go back to summary

- ~90% - determined by macromolecular crystallography
- How do I assess the quality of structure?

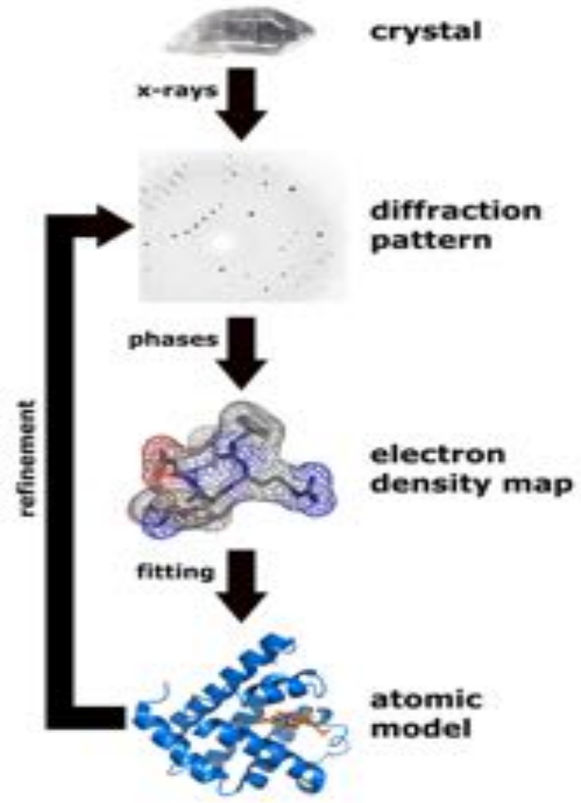


Image by Thomas Splettstoesser (www.scistyle.com)

Experimental Details Hide

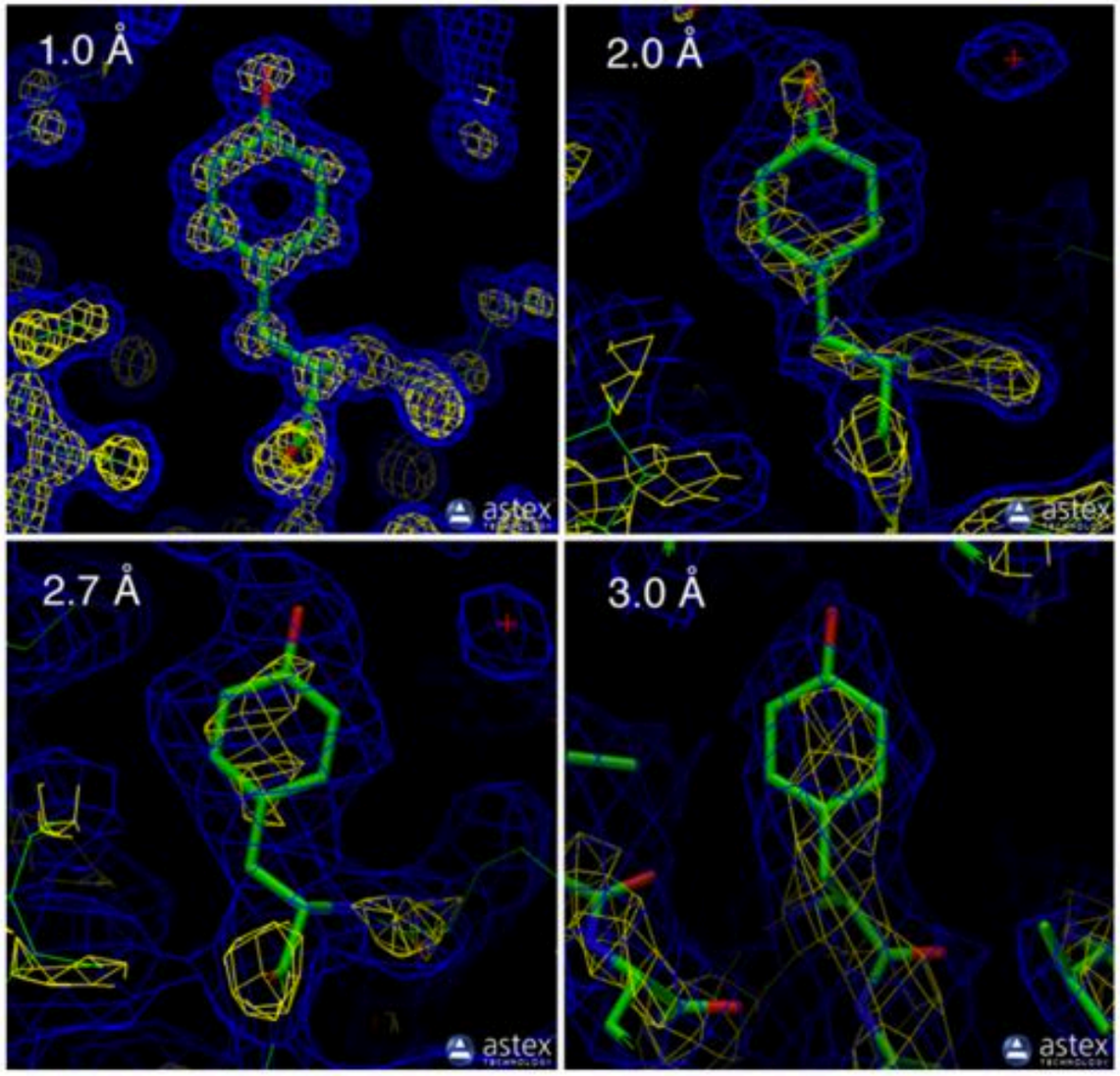
Method: X-RAY DIFFRACTION
Exp. Data:
BMRB
Structure Factors
EDS

Resolution[Å]:	1.80
R-Value:	0.176 (obs.)
R-Free:	n/a
Space Group:	$P\ 2_1\ 2_1\ 2_1$

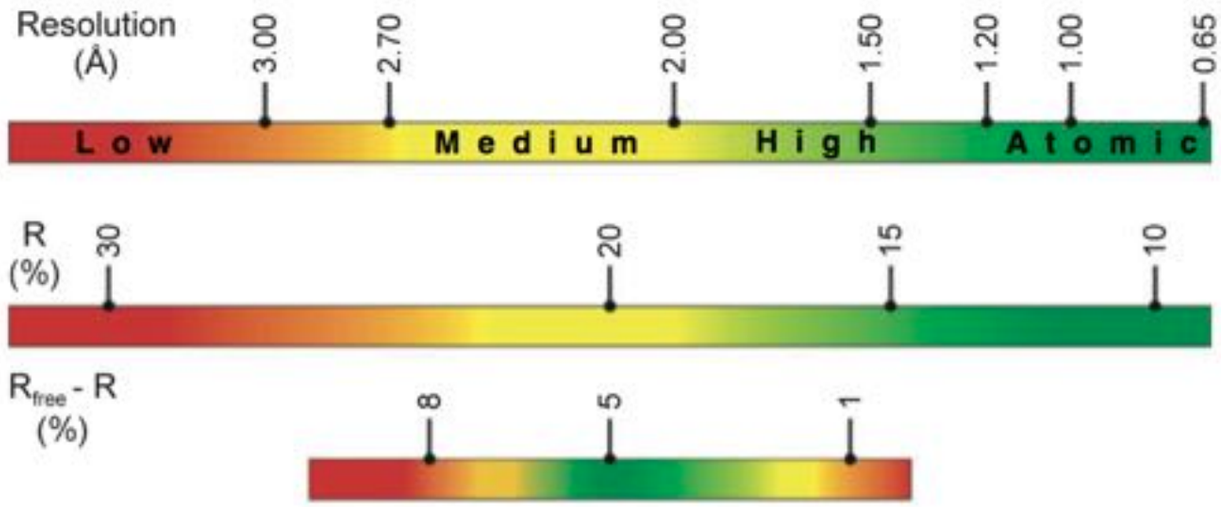
Unit Cell:

Length [Å]	Angles [°]
a = 50.84	α = 90.00
b = 42.77	β = 90.00
c = 28.95	γ = 90.00

Resolution, R-factor and quality



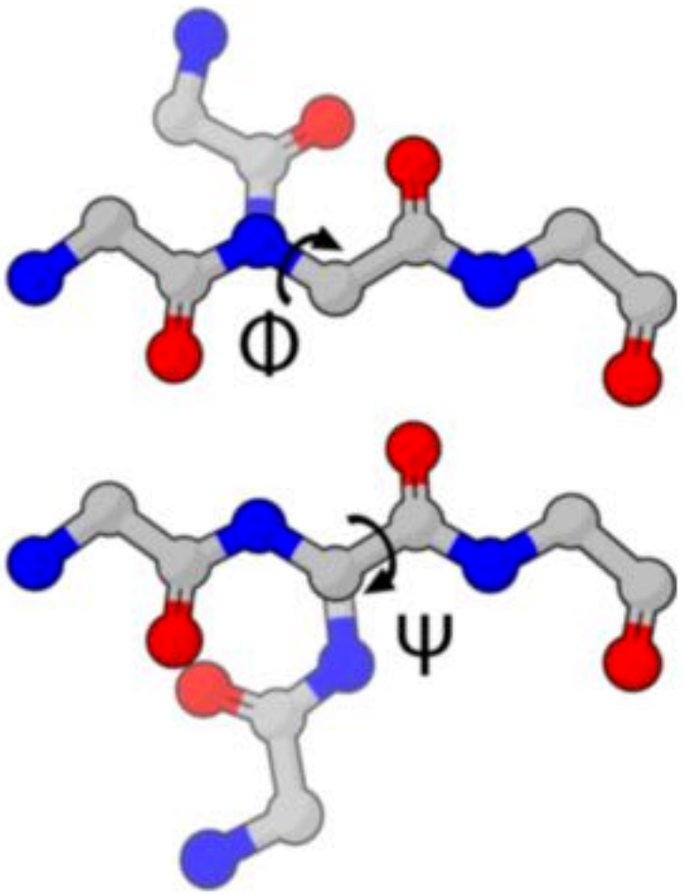
Electron density maps at different resolutions

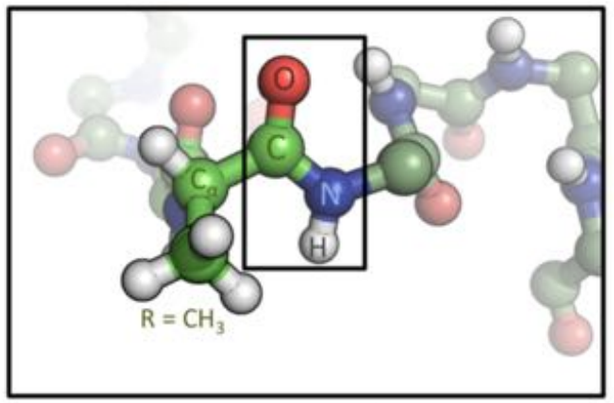


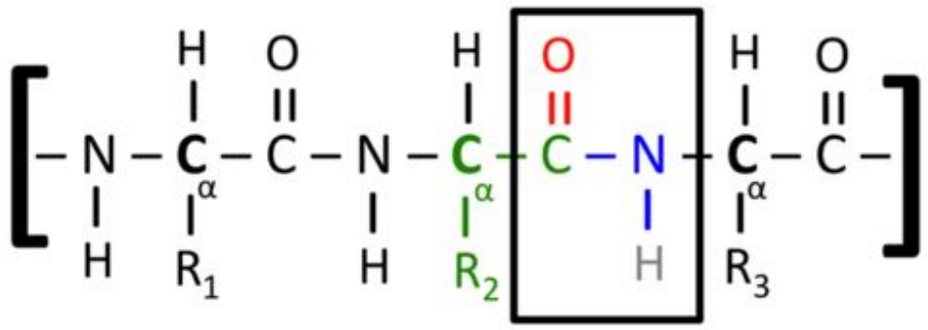
From Wlodawer A, Minor W, Dauter Z, Jaskolski M. FEBS J. 2008 Jan;275(1):1-21

From http://www.rcsb.org/pdb/101/static101.do?p=education_discussion/Looking-

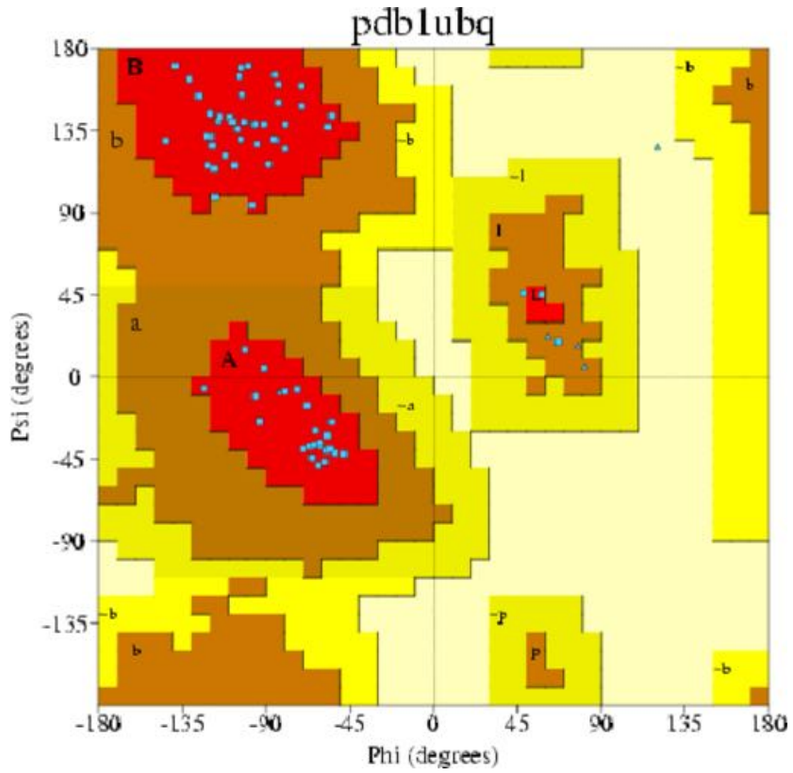
Checking Ramachandran plot







- On PDB page click Links menu
- Click PDBSum
- Click PROCHECK



PDB format

- Return back: You can Google 1ubq
- Click Display file -> PDB File

B-factor

Chain ID

```
ATOM      1  N   MET A   1      27.340  24.430   2.614  1.00  9.67           N
ATOM      2  CA  MET A   1      26.266  25.413   2.842  1.00 10.38           C
ATOM      3  C   MET A   1      26.913  26.639   3.531  1.00  9.62           C
ATOM      4  O   MET A   1      27.886  26.463   4.263  1.00  9.62           O
ATOM      5  CB  MET A   1      25.112  24.880   3.649  1.00 13.77           C
ATOM      6  CG  MET A   1      25.353  24.860   5.134  1.00 16.29           C
ATOM      7  SD  MET A   1      23.930  23.959   5.904  1.00 17.17           S
ATOM      8  CE  MET A   1      24.447  23.984   7.620  1.00 16.11           C
ATOM      9  N   GLN A   2      26.335  27.770   3.258  1.00  9.27           N
ATOM     10  CA  GLN A   2      26.850  29.021   3.898  1.00  9.07           C
```

Atom name

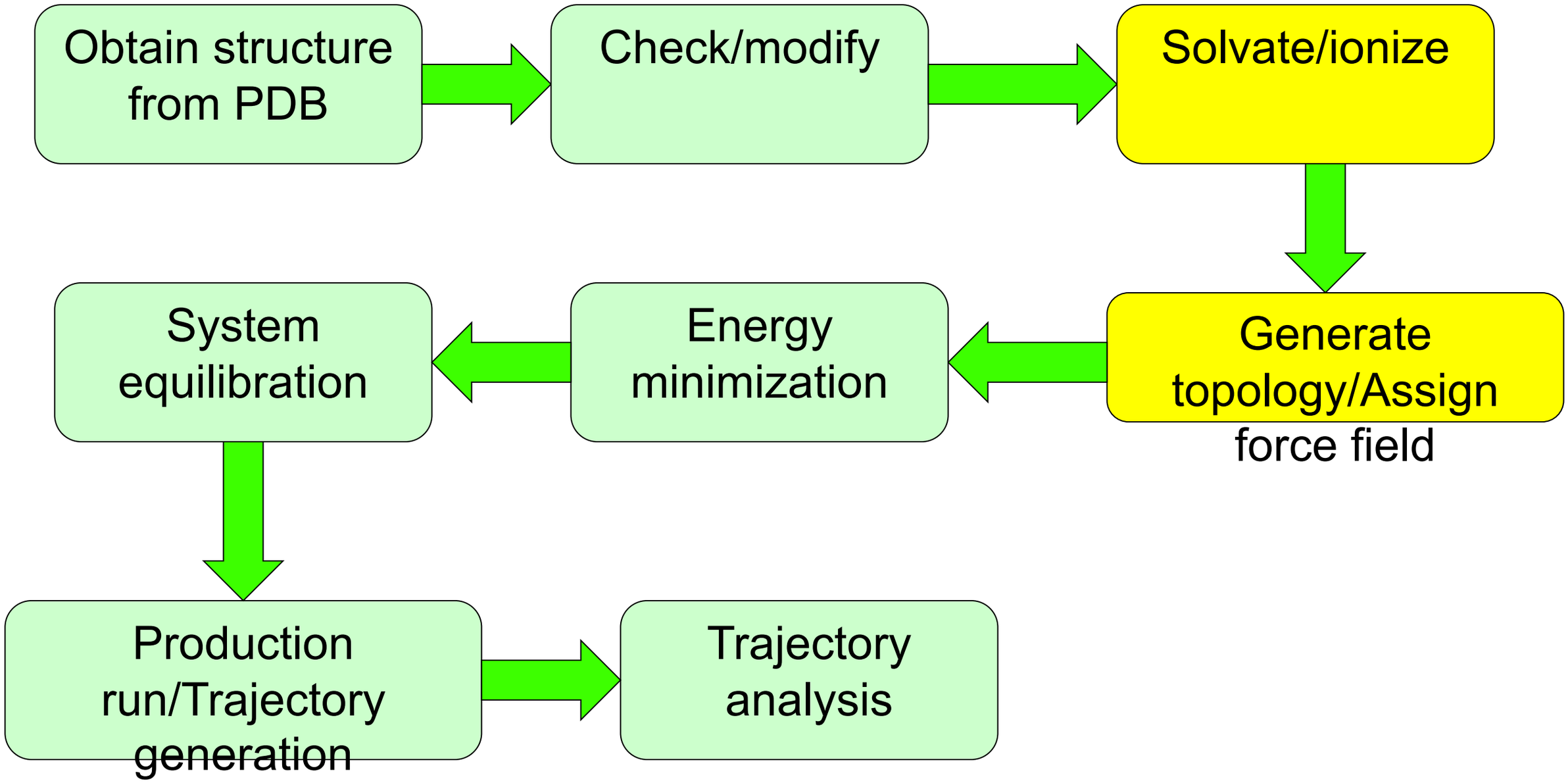
Residue name

Residue number

X,Y,Z

Occupancy

How do I do an MD simulation?



Topology file

```
RESI ARG          1.00
GROUP
ATOM N    NH1    -0.47  !      |                        HH11
ATOM HN   H       0.31  !   HN-N                         |
ATOM CA   CT1     0.07  !      |    HB1 HG1 HD1 HE     NH1-HH12
ATOM HA   HB      0.09  !      |    |   |   |   |    //(+)
GROUP                   !   HA-CA--CB--CG--CD--NE--CZ
ATOM CB   CT2    -0.18  !      |    |   |   |          \
ATOM HB1  HA      0.09  !      |    HB2 HG2 HD2         NH2-HH22
ATOM HB2  HA      0.09  !    O=C                          |
GROUP                   !      |                        HH21
ATOM CG   CT2    -0.18
ATOM HG1  HA      0.09
ATOM HG2  HA      0.09
GROUP
ATOM CD   CT2     0.20
ATOM HD1  HA      0.09
ATOM HD2  HA      0.09
ATOM NE   NC2    -0.70
ATOM HE   HC      0.44
ATOM CZ   C       0.64
ATOM NH1  NC2    -0.80
ATOM HH11 HC      0.46
```

Force field parameters file

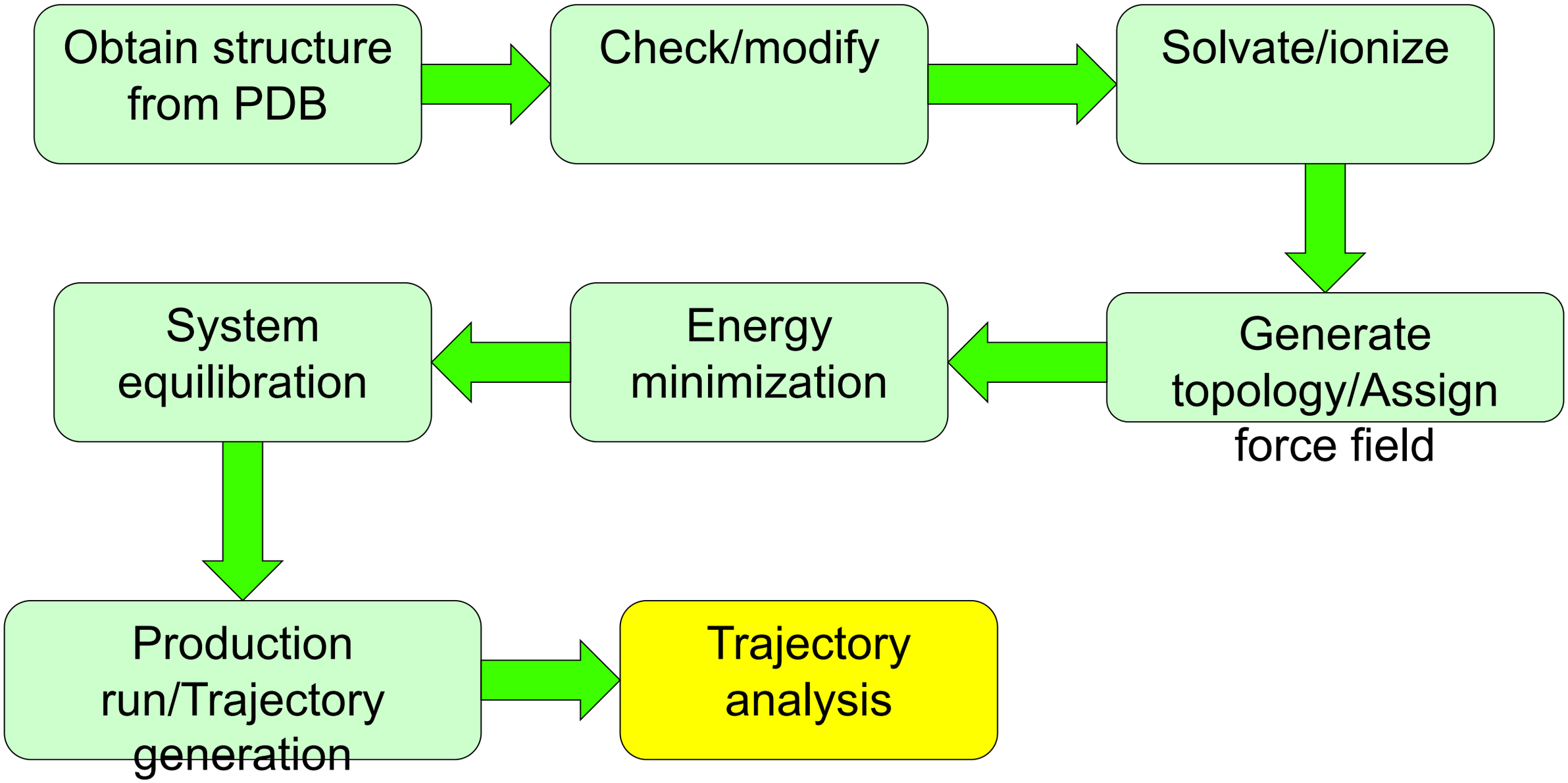
```
BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type Kb          b0
!
!Carbon Dioxide
CST  OST   937.96      1.1600 ! JES
!Heme to Sulfate (PSUL) link
SS   FE    250.0       2.3200 !force constant a guess
           !equilbrium bond length optimized to reproduce
           !CSD survey values of
           !2.341pm0.01 (mean, standard error)
           !adm jr., 7/01
C    C     600.000     1.3350 ! ALLOW ARO HEM
                ! Heme vinyl substituent (KK, from propene (JCS))
CA   CA    305.000     1.3750 ! ALLOW   ARO
                ! benzene, JES 8/25/89
CE1  CE1   440.000     1.3400   !
          ! for butene; from propene, yin/adm jr., 12/95
CE1  CE2   500.000     1.3420   !
```

NAMD conf file

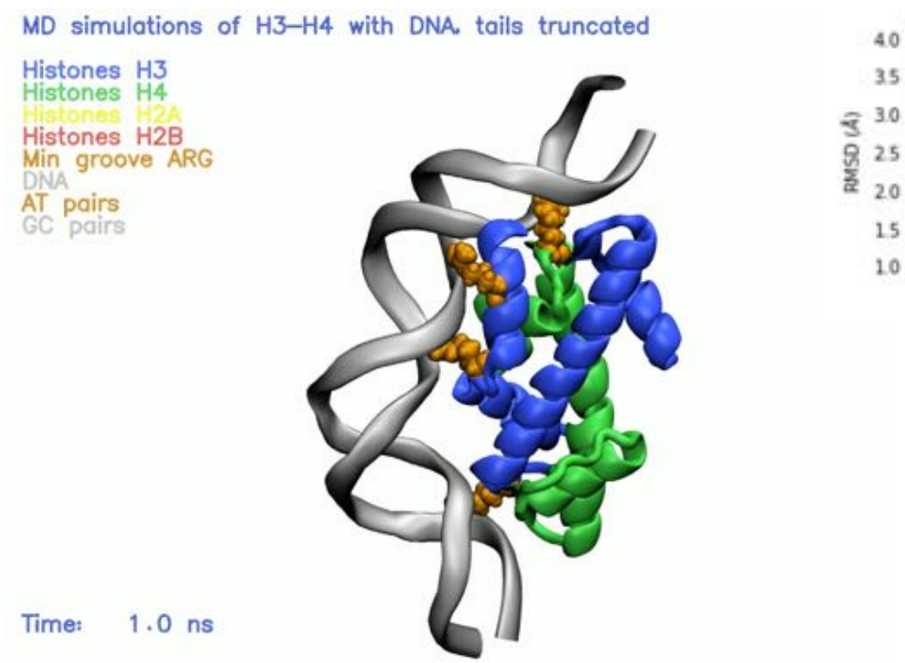
```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################

# Minimization and simulation
# of Ubiquitin in a Water box


#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure          ubq_ready.psf
coordinates        ubq_ready.pdb

set temperature    310
set outputname     simul

firsttimestep      0


#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################
```

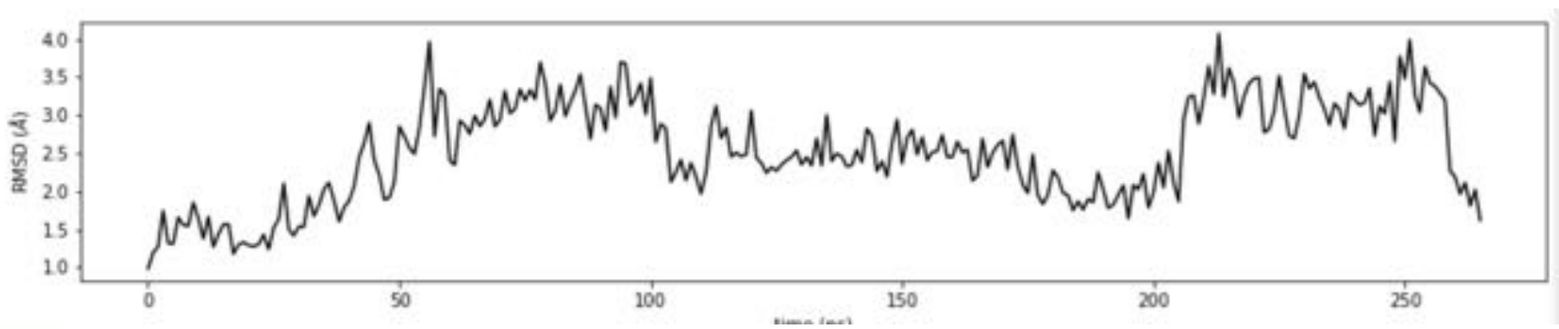
.....

How do I do an MD simulation?

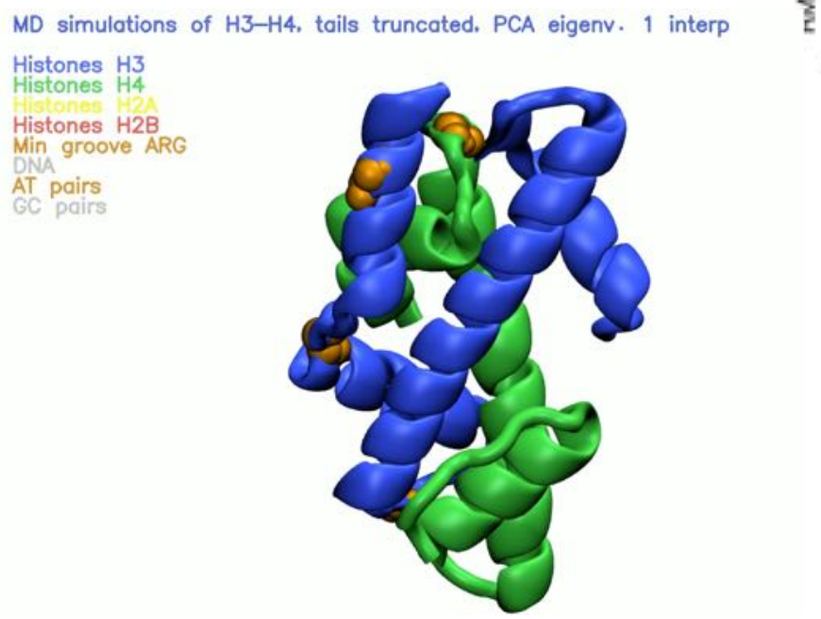


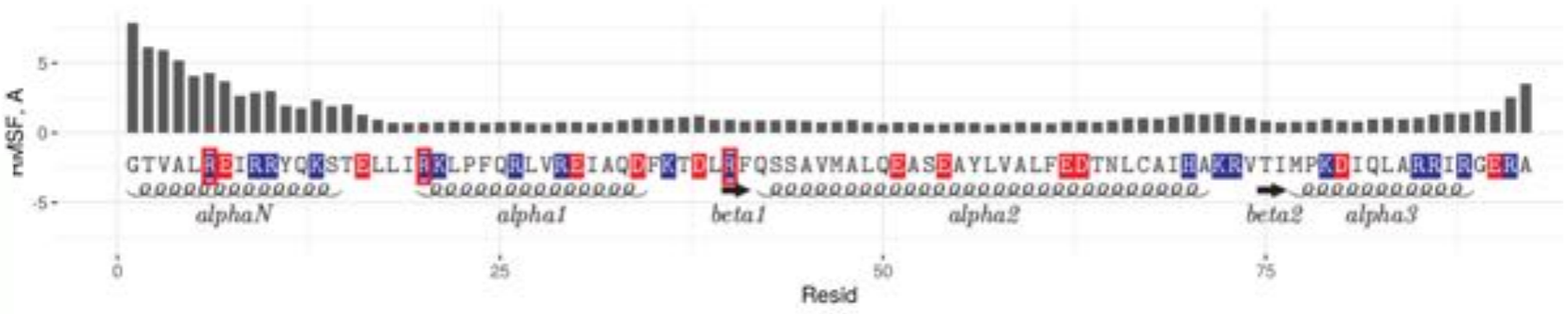
Trajectory analysis





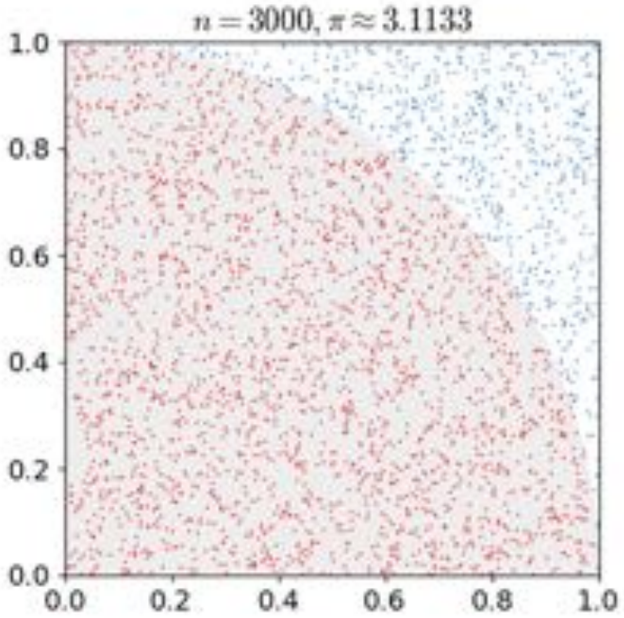
$$RMSD = \min_{rot,trans} \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\vec{r_a} - \vec{r_{a0}})^2}$$





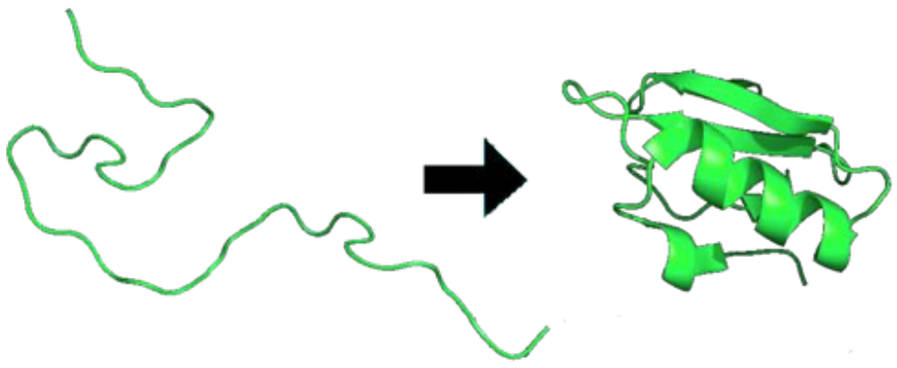
Методы Монте-Карло

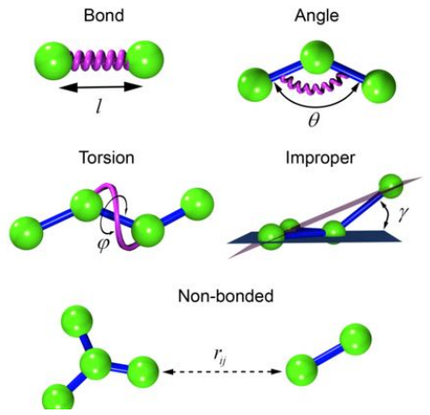
- Решение детерминированных задач с использованием генератора случайных числе.



Вычисление числа Пи

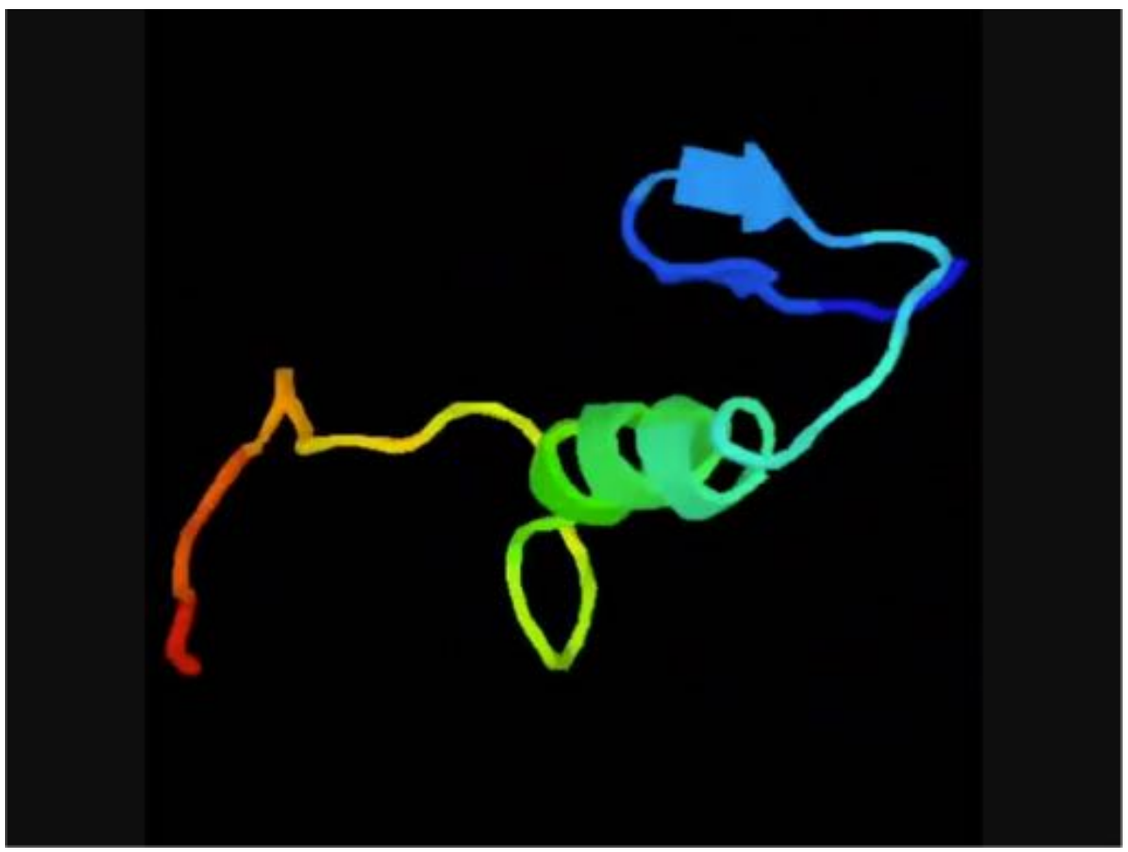
Распределение Гиббса/Больцмана





$$U(\{\vec{r}_i\}) = \sum_{bonds} \frac{1}{2} k_b (l - l_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n [1 + \cos(n\varphi - \varphi_0)]$$

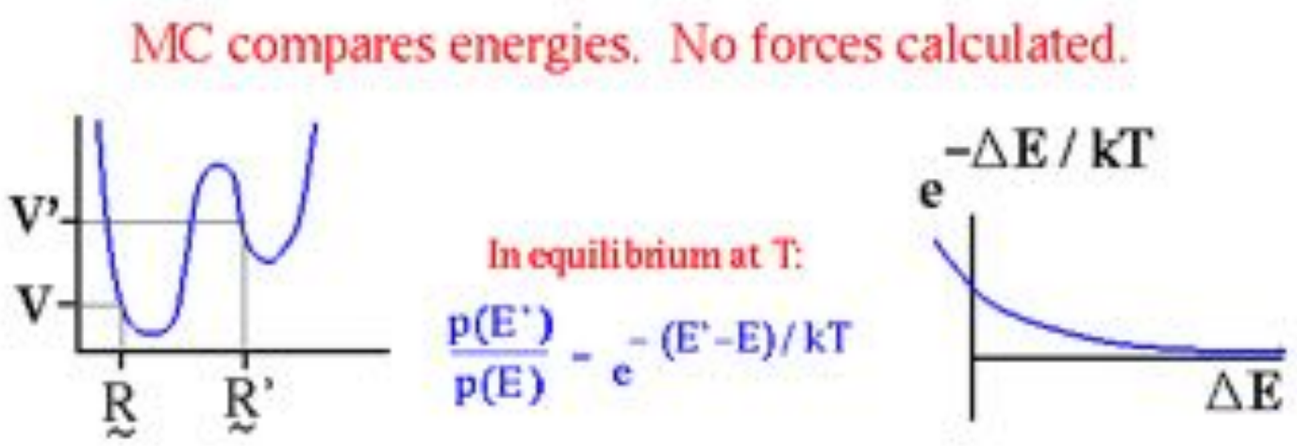
$$+ \sum_{impropers} \frac{1}{2} k_\gamma (\gamma - \gamma_0)^2 + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} f_{ij}$$



$$P_{microstate} \propto e^{-\frac{U}{kT}}$$

Распределение Гиббса

Алгоритм Метрополиса-Хастингса

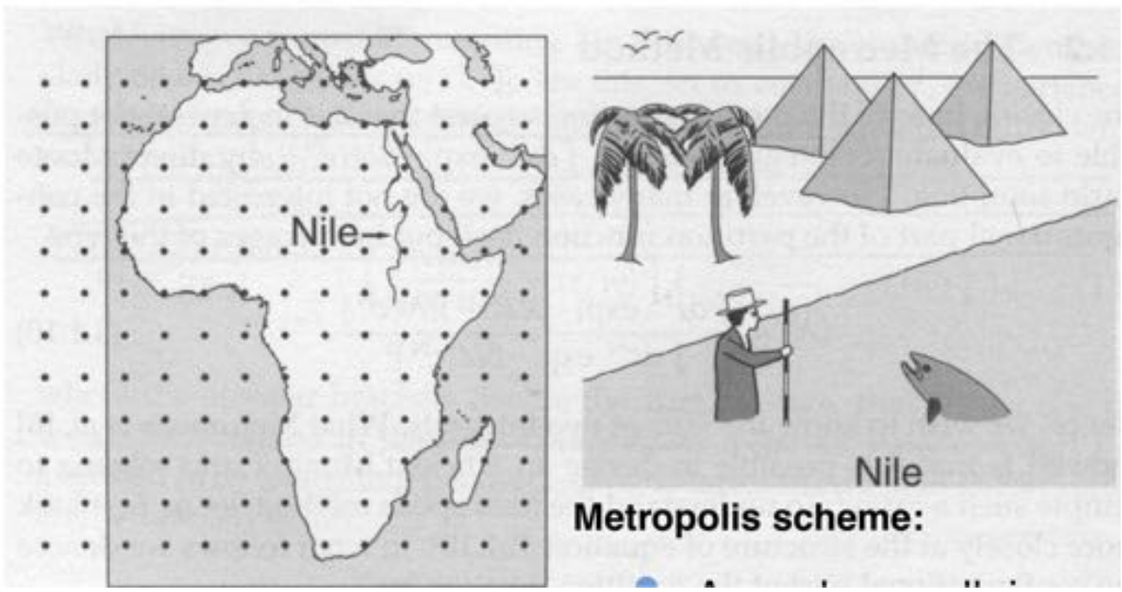


$$P_{microstate} \propto e^{-\frac{U}{kT}}$$

Gibbs distribution

- Prescribes transition probabilities to satisfy detailed balance, given desired limiting distribution
- Recipe:
  From a state *i*…
  - *with probability* τ_{ij}, choose a trial state *j* for the move (note: $\tau_{ij} = \tau_{ji}$)
  - *if* $\pi_j > \pi_i$, *accept* j *as the new state*
  - *otherwise, accept state* j *with probability* π_j/π_i

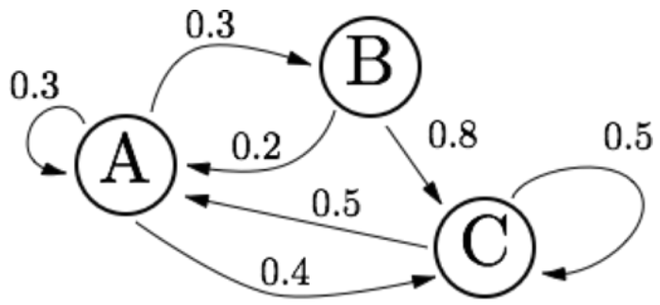
    generate a random number R on (0,1); accept if $R < \pi_j/\pi_i$
  - *if not accepting* j *as the new state, take the present state as the next one in the Markov chain* ($\pi_{ii} \neq 0$)



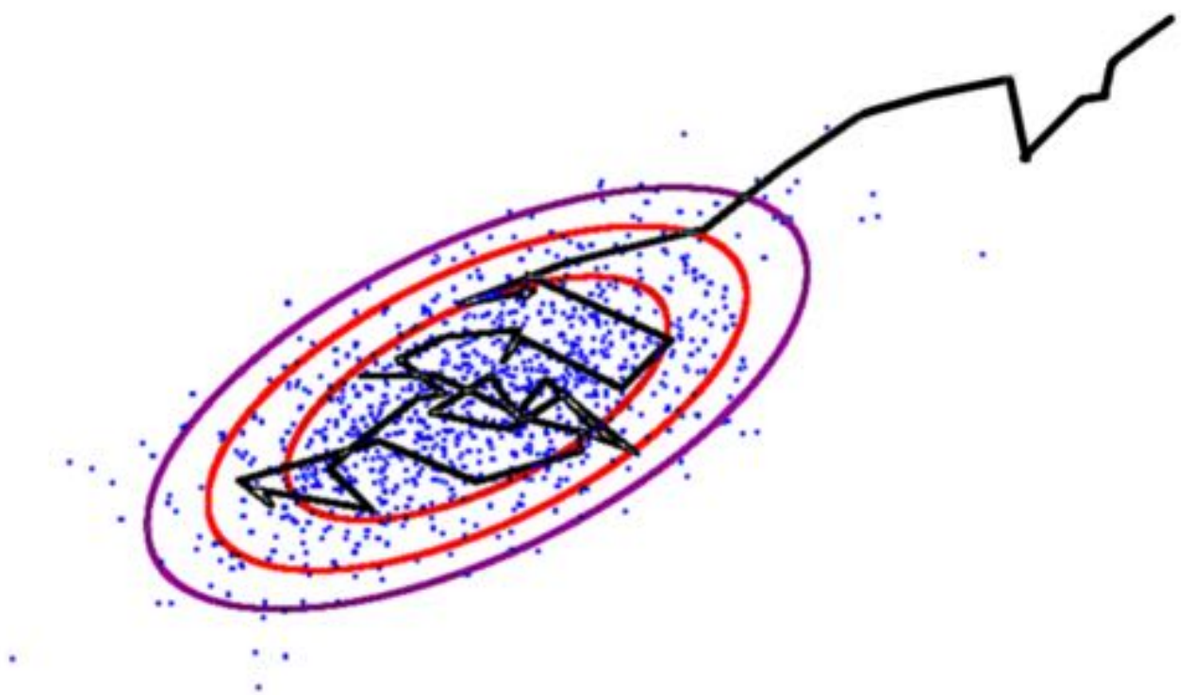
Metropolis scheme:

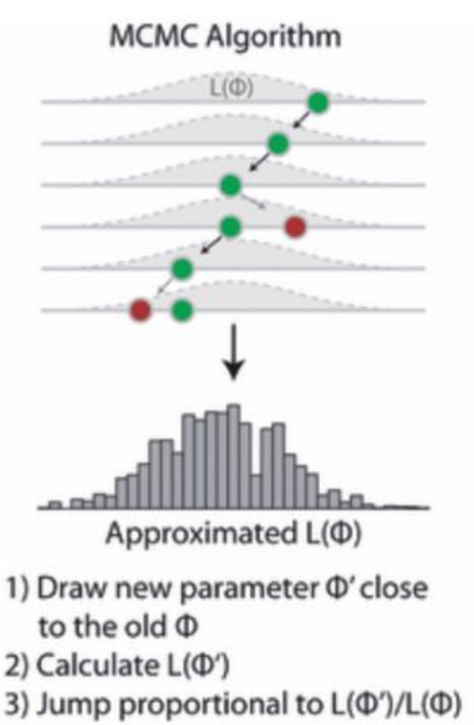
Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, J. Chem. Phys., 21 1087 (1953)

Markov Chain Monte Carlo (MCMC)

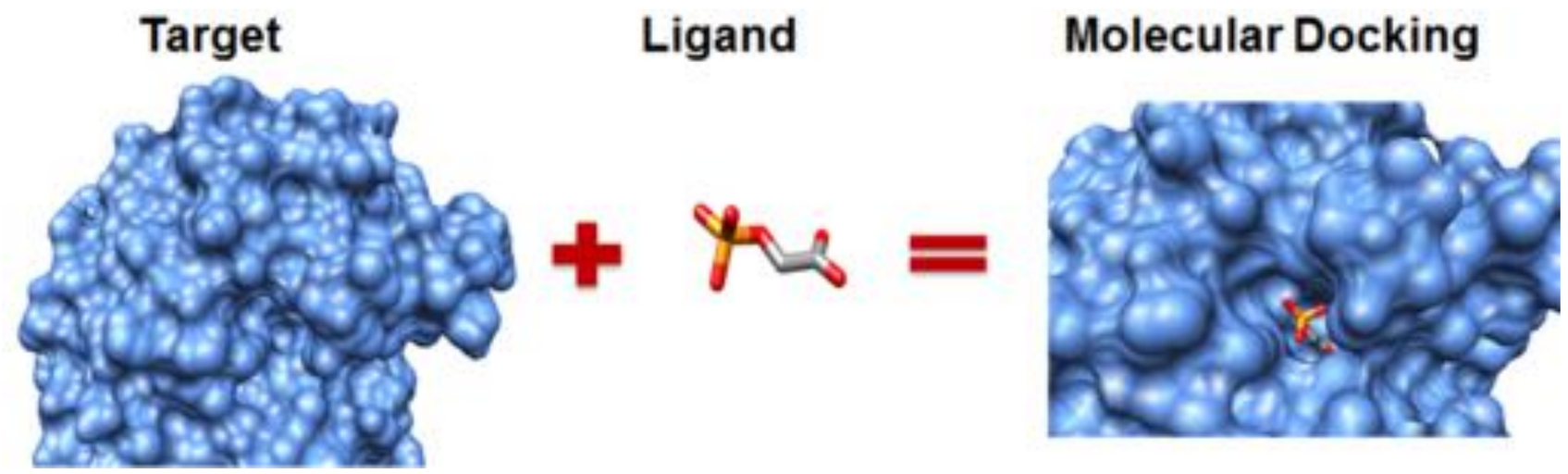


Цепи Маркова

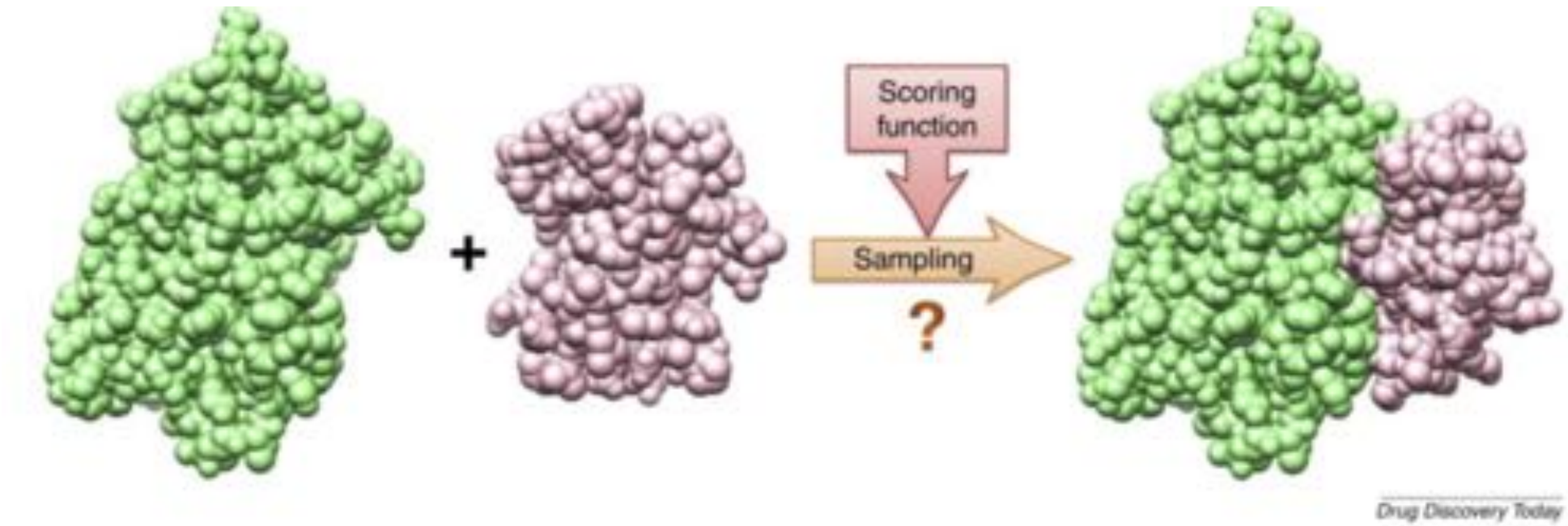




Методы докинга

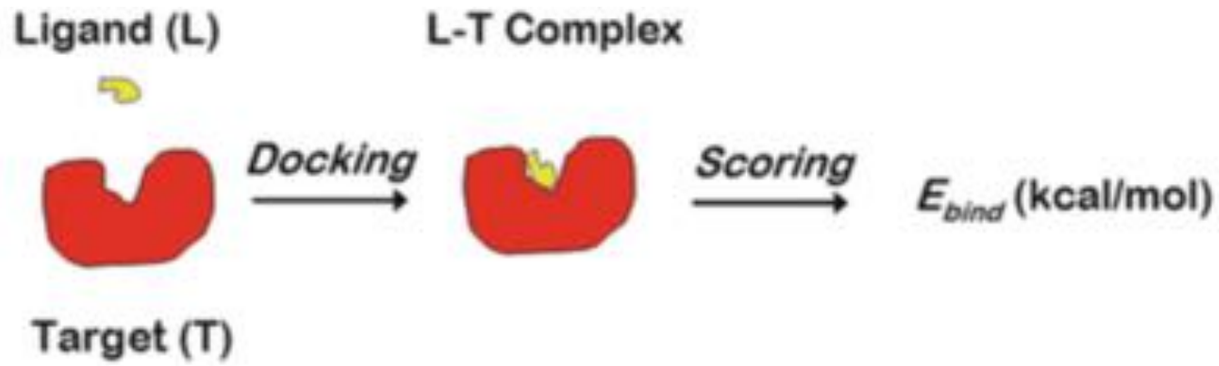


Докинг лигандов



Белок-белковый докинг

Методы докинга: функции скоринга



1. Molecular mechanics force field-based scoring function:

$$E_{bind} = \sum_{i=1}^{lig}\sum_{j=1}^{rec}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332\frac{q_i q_j}{\varepsilon_0 r_{ij}}\right)$$

2. Empirical scoring function:

$$E_{bind} = E_0 + E_{hb}\sum_{hb} f(\Delta R, \Delta\alpha) + E_{ionic}\sum_{ionic} f(\Delta R, \Delta\alpha) + E_{lipo}\sum_{lipo}\left|A_{lipo}\right| + E_{rot}NROT$$

3. Knowledge-based scoring function:

$$E_{bind} = \gamma\sum_{i=1}^{lig}\sum_{j=1}^{rec}E_{ij}(r) + (1-\gamma)\times\left[\sum_{i=1}^{lig}E_i(SASA, SASA_0) + \sum_{j=1}^{rec}E_j(SASA, SASA_0)\right]$$

История и некоторые примеры

- **1953** – Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. "Equations of State Calculations by Fast Computing Machines", JCC, **Monte-Carlo Method used for particle simulations**

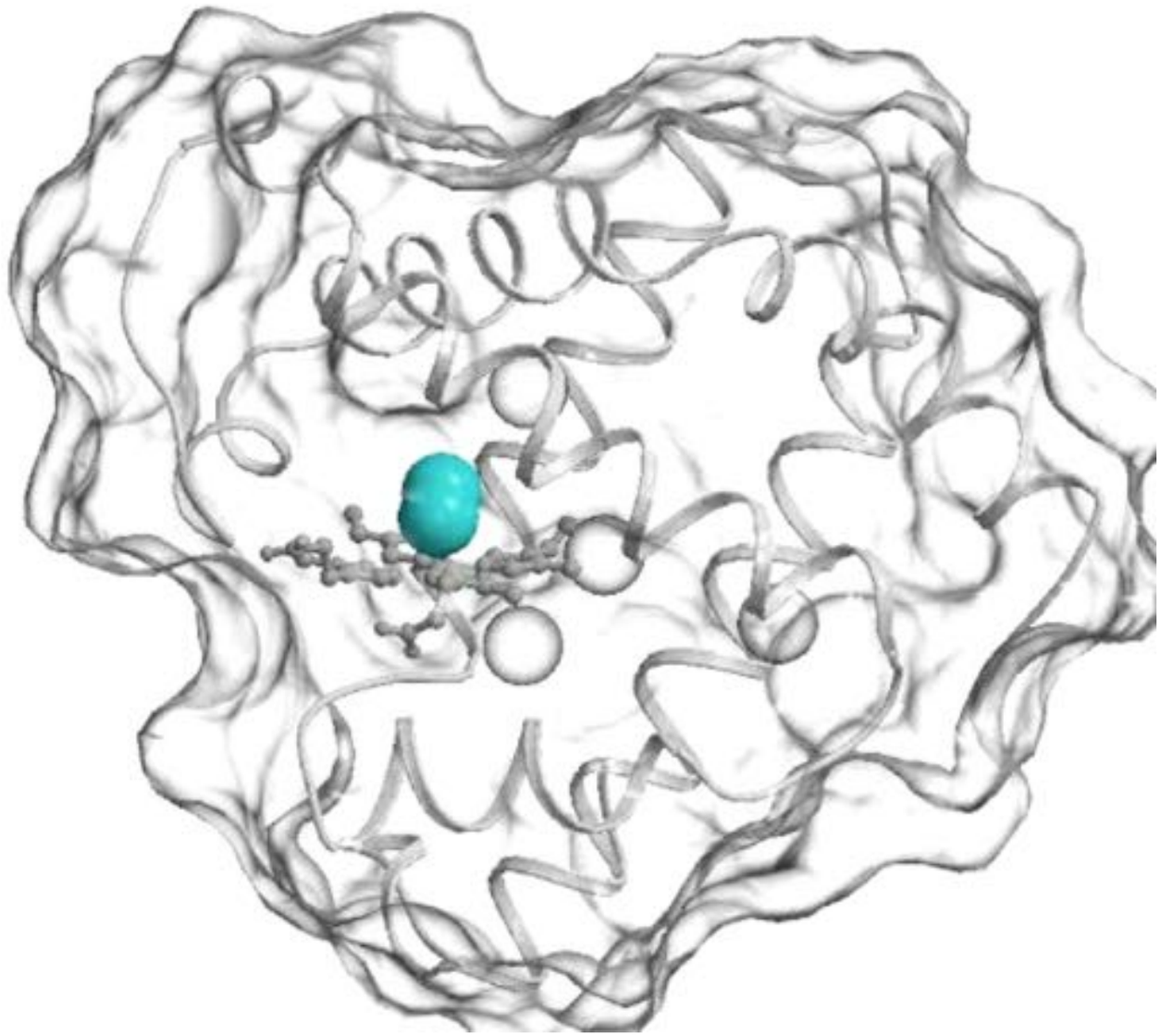
- **1957** - Alder, B. J.; T. E. Wainwright (1959). "Studies in Molecular Dynamics. I. General Method". *J. Chem. Phys.* **31** (2): 459 – **First MD simulations**

- **1975** - M. Levitt & A. Warshel, "**Computer simulation of protein folding",** Nature

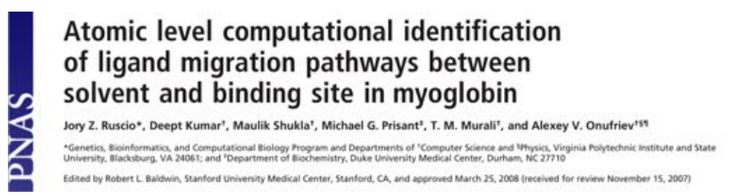
- **1979 - Dynamics of ligand binding to heme protein,** DA Case, M Karplus *J Mol Biol*

- **1983** - Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983). "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations". *J Comp Chem* **4** (2): 187–217.
- **1988** - M Levitt and R Sharon "Accurate simulation of protein dynamics in solution" **First MD in explicit solvent**

История и некоторые примеры



Диффузия CO через порталы в миоглобине



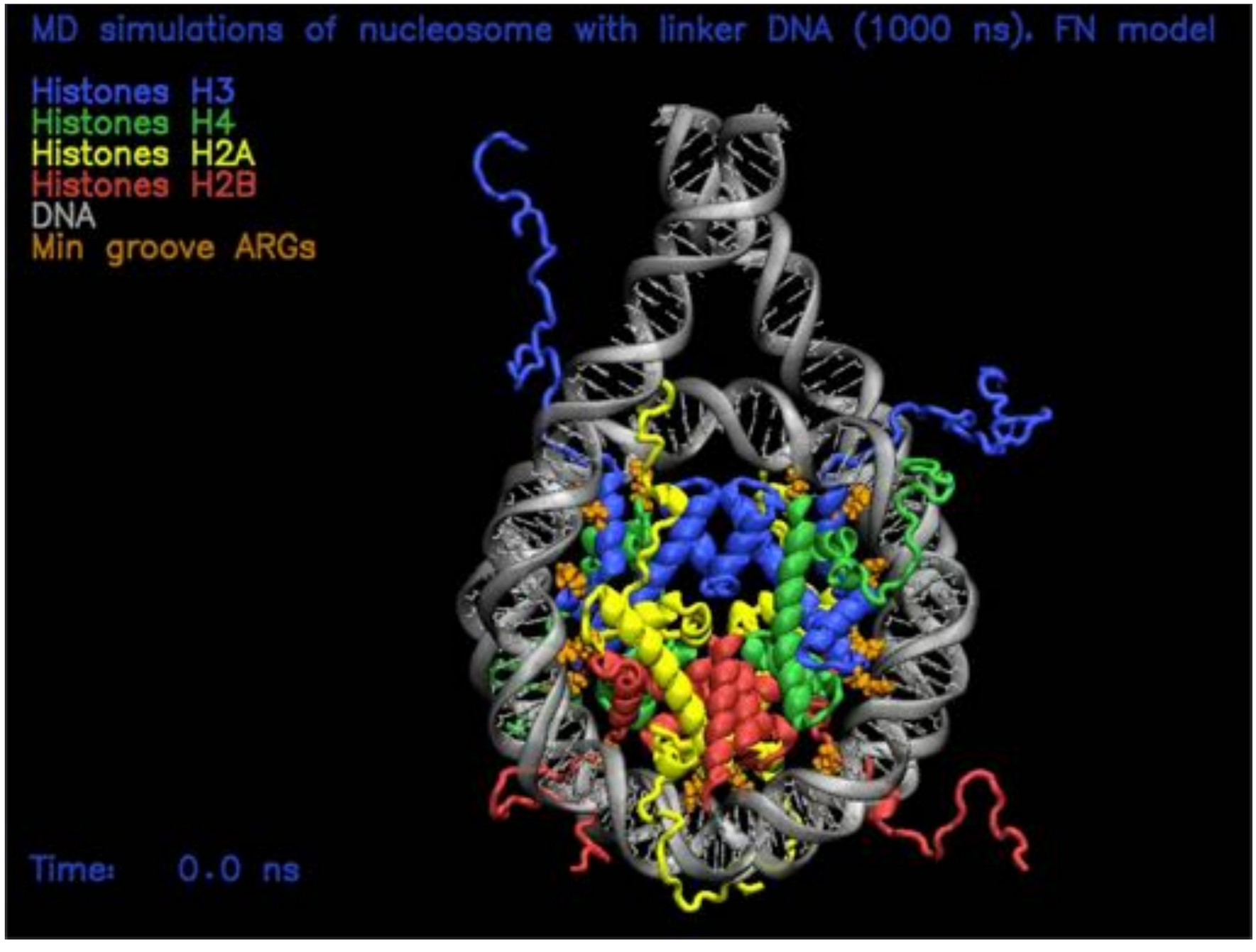
Atomic level computational identification of ligand migration pathways between solvent and binding site in myoglobin

Jory Z. Ruscio*, Deept Kumar†, Maulik Shukla†, Michael G. Prisant‡, T. M. Murali†, and Alexey V. Onufriev†§¶

*Genetics, Bioinformatics, and Computational Biology Program and Departments of †Computer Science and §Physics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061; and ‡Department of Biochemistry, Duke University Medical Center, Durham, NC 27710

Edited by Robert L. Baldwin, Stanford University Medical Center, Stanford, CA, and approved March 25, 2008 (received for review November 15, 2007)

История и некоторые примеры



Нуклеосомы

jmb
Journal of Molecular Biology
Volume 428, Issue 1, 16 January 2016, Pages 221-237

Coupling between Histone Conformations and DNA Geometry in Nucleosomes on a Microsecond Timescale: Atomistic Insights into Nucleosome Functions

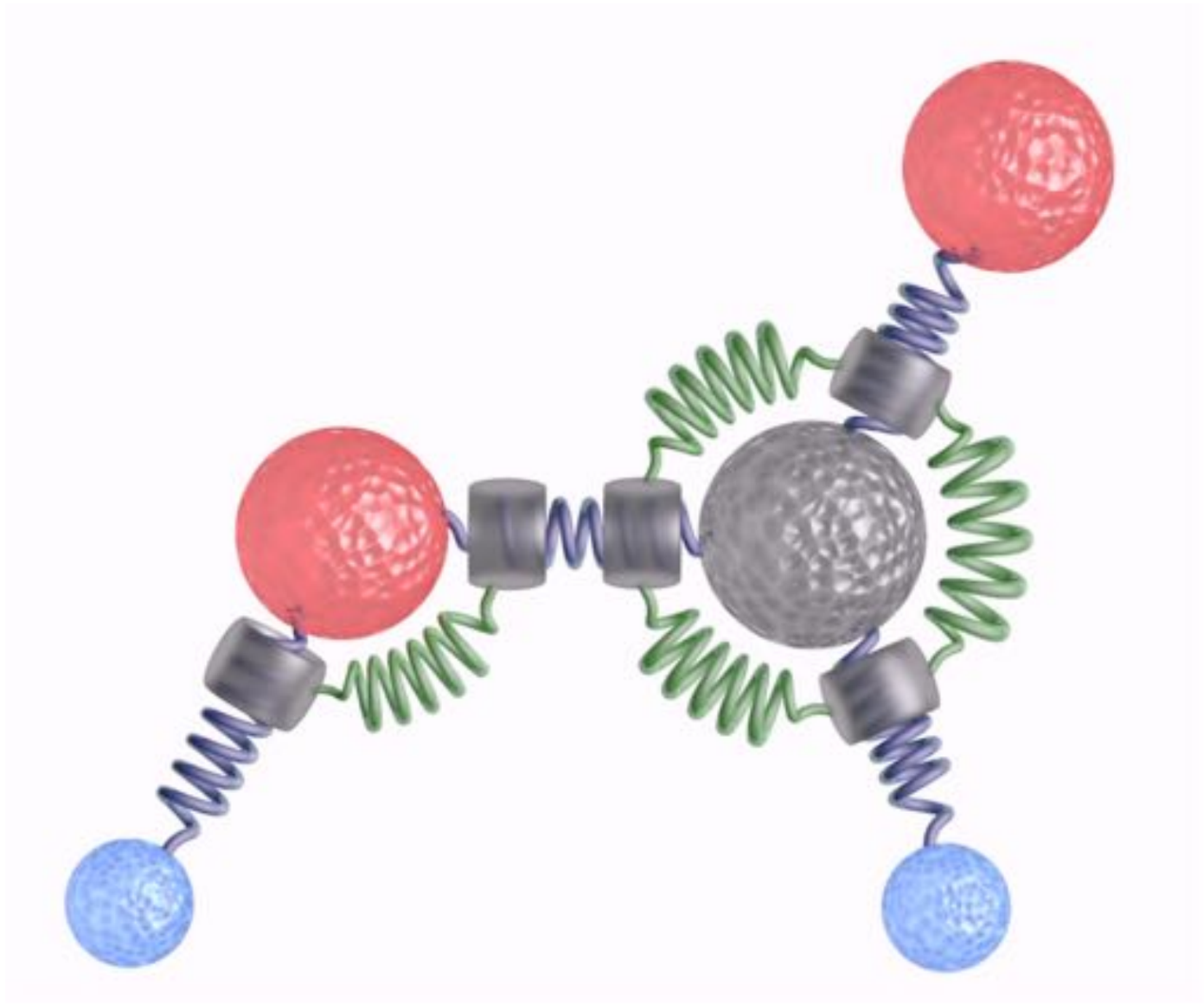
Alexey K. Shaytan [1,2], Grigoriy A. Armeev [2], Alexander Goncearenco [3], Victor B. Zhurkin [3], David Landsman [1], Anna R. Panchenko [1]

История и некоторые примеры

Ионные каналы

Sauguet L*, Poitevin F*, Murail S*, Van Renterghem C, Moraga-Cid G, Malherbe L, Thompson AW, Koehl P, Corringer PJ, Baaden M, Delarue M. Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels. EMBO J. 32, 728-41. *Contributed equally.
emboj.embopress.org/content/32/5/728

Thanks for your attention!



Suggested reading:

Text books:

•D. Frenkel, B. Smit, "Understanding Molecular Simulation, Second Edition: From Algorithms to Applications"

Reviews:

•Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE, "Biomolecular simulation: a computational microscope for molecular biology", Annu Rev Biophys. 2012;41:429-52

Useful papers:

•Wlodawer A, Minor W, Dauter Z, Jaskolski M., "Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures", FEBS J. 2008 Jan;275(1):1-21

Web links:

•NAMD http://www.ks.uiuc.edu/Research/namd/
•VMD http://www.ks.uiuc.edu/Research/vmd/
•AMBER http://ambermd.org
•CHARMM http://www.charmmtutorial.org
•GROMACS http://www.gromacs.org
•LAMMPS http://lammps.sandia.gov